```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:15:33 ; Search time 9.71143 Seconds
                                           (without alignments)
                                           2813.016 Million cell updates/sec

Title:           US-10-006-011A-9
Perfect score:   1566
Sequence:        1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        PIR_78:*
               1:  pir1:*
               2:  pir2:*
               3:  pir3:*
               4:  pir4:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.

                                    SUMMARIES
                        %
Result               Query
  No.       Score    Match Length DB  ID                           Description
------------------------------------------------------------------------------------
     1       1566    100.0   4391  2  A38096                       perlecan precursor
     2     1370.5     87.5   3707  2  S18252                       heparan sulfate pr
     3        349     22.3   3375  2  T19821                       hypothetical prote
     4      333.5     21.3   1959  1  AGRT                         agrin - rat
     5      314.5     20.1   1328  2  T43060                       agrin - electric r
     6      301.5     19.3   1955  1  AGCH                         agrin precursor -
     7      283.5     18.1   1531  2  T42218                       slit-1 protein hom
     8        264     16.9   4307  2  T20721                       hypothetical prote
     9        264     16.9   5147  1  IJFFTM                       cadherin-related t
    10        249     15.9    601  2  T22025                       hypothetical prote
    11        249     15.9    601  2  D89711                       protein F40E10.4 [
    12        246     15.7   1523  2  T13953                       MEGF5 protein - ra
    13      243.5     15.5   1025  2  T42626                       secreted leucine-r
    14      223.5     14.3   1751  1  MMHUMH                       laminin alpha-2 ch
    15        217     13.9   1438  2  A48216                       neurexin III-alpha
    16        217     13.9   1471  2  B48218                       neurexin III-alpha
    17        217     13.9   1578  2  I48216                       neurexin III-alpha
    18        214     13.7   2809  2  T30213                       G-cadherin - sea u
    19      207.5     13.3   3106  1  S53868                       laminin alpha-2 ch
    20        199     12.7   2321  2  S78549                       notch3 protein - h
    21      197.5     12.6   1530  2  I45944                       neurexin I-alpha -
    22      197.5     12.6   2610  2  T20968                       hypothetical prote
    23      196.5     12.5   1507  2  A40228                       neurexin I-alpha p
    24      194.5     12.4   1715  2  C40228                       neurexin II-alpha
    25        191     12.2   3034  2  T14119                       seven-pass transme
    26      188.5     12.0   3075  2  S14458                       laminin alpha-1 ch
    27      185.5     11.8   3672  2  T23433                       hypothetical prote
    28      185.5     11.8   3704  2  T37316                       probable laminin a
    29        184     11.7   3097  2  T00021                       DN-cadherin - frui
```

```
    30        183     11.7   1469  2  B36665                       slit protein 2 pre
    31        183     11.7   1480  2  A36665                       slit protein 1 pre
    32        182     11.6   2318  2  S45306                       notch 3 protein -
    33      181.5     11.6   2531  2  S18188                       notch protein homo
    34      181.5     11.6   2531  2  A46019                       notch-1 protein -
    35        180     11.5   2531  2  T31070                       notch homolog - se
    36      177.5     11.3   1964  2  T09059                       notch4 - mouse
    37      176.5     11.3    473  2  A56175                       adhesive plaque pr
    38        176     11.2   1647  2  T32934                       hypothetical prote
    39        175     11.2   1203  2  A49175                       Motch B protein -
    40        175     11.2   2471  2  A49128                       cell-fate determin
    41      174.5     11.1   2555  2  A40043                       notch protein homo
    42        173     11.0   1220  2  A56136                       jagged protein pre
    43        170     10.9    385  2  A54785                       preadipocyte facto
    44        170     10.9    385  2  S53718                       homeotic protein d
    45      169.5     10.8   1268  2  S52781                       neurocan - mouse

                                    ALIGNMENTS

RESULT 1
A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prot
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515; NID:g29469; PIDN:CAA44373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989
```

```
A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
anes.
A;Reference number: A33625; MUID:90078352; PMID:2687294
A;Accession: B33625
A;Molecule type: protein
A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A;Accession: A33625
A;Molecule type: protein
A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A;Note: peptide potentially matches four different regions of sequence shown
C;Genetics:
A;Gene: GDB:HSPG2
A;Cross-references: GDB:126372; OMIM:142461
A;Map position: 1p36.1-1p36.1
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-4391/Product: perlecan #status predicted <MAT>
F;22-193/Domain: I <DOM1>
F;194-530/Domain: II <DOM2>
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;531-1676/Domain: III <DOM3>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;1677-3686/Domain: IV <DOM4>
F;2007-2034/Domain: transmembrane #status predicted <TRM>
F;3687-4391/Domain: V <DOM5>
F;3845-3880/Domain: EGF homology <EGF1>
F;3888-3921/Domain: EGF homology <EGF>
F;3953-4106/Domain: laminin G repeat homology <LG2>
F;4147-4175/Domain: EGF homology <EGF2>
F;4149-4151/Region: motor neuron attachment (L-R-E) motif
F;4299-4301/Region: motor neuron attachment (L-R-E) motif
F;65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval
F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

  Query Match             100.0%;  Score 1566;  DB 2;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-105;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4108 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4167

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4168 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4227

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4228 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 4287

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4288 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 4347

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||||||||||||||||||
Db       4348 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
S18252
heparan sulfate proteoglycan - mouse
N;Alternate names: perlecan
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S18252; A31917; B31917; S66460
```

```
R;Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteog]
adhesion molecule.
A;Reference number: S18252; MUID:92078153; PMID:1744087
A;Accession: S18252
A;Molecule type: mRNA
A;Residues: 1-3707 <NOO>
A;Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hase
J. Biol. Chem. 263, 16379-16387, 1988
A;Title: Identification of cDNA clones encoding different domains of the basement membra
A;Reference number: A92680; MUID:89034110; PMID:2972708
A;Accession: A31917
A;Molecule type: mRNA
A;Residues: 940-1601 <NO2>
A;Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
A;Accession: B31917
A;Molecule type: mRNA
A;Residues: 1870-2600 <NO3>
A;Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A;Reference number: S66460; MUID:95377282; PMID:7649154
A;Accession: S66460
A;Molecule type: protein
A;Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: glycoprotein
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;764-811/Domain: laminin-type EGF-like homology <LEG>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;3163-3198/Domain: EGF homology <EGF>
F;3270-3423/Domain: laminin G repeat homology <LG2>
F;3464-3492/Domain: EGF homology <EGF7>
F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             87.5%;  Score 1370.5;  DB 2;  Length 3707;
  Best Local Similarity   86.3%;  Pred. No. 4e-91;
  Matches  245;  Conservative   20;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||::||||||||||||||||:||||||||||||||||| ||||:||||:| |||||
Db       3425 CERQPCRNGATCMPAGEYEFQCLCQDGFKGDLCEHEENPCQLHEPCLNGGTCRGARCLCL 3484

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              |||||||||||:|:|: ||||| ||||||||||||||||:|:|||  ||: |||||||||
Db       3485 PGFSGPRCQQGAGYGVVESDWHPEGSGGNDAPGQYGAYFYDNGFLGLPGNSFSRSLPEVP 3544

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              |||| ||||||| |||||||| | || : ||||||||||||||| |||||||||||| ||
Db       3545 ETIEFEVRTSTADGLLLWQGV-VREASRSKDFISLGLQDGHLVFSYQLGSGEARLVSGDP 3603

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||:||||||:|||||||||:||:||||||||||| |  :|||||||||||| |:
Db       3604 INDGEWHRITALREGQRGSIQVDGEDLVTGRSPGPNVAVNTKDIIYIGGAPDVATLTRGK 3663

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||:||||||:||||||||||||||||||||||||||||
Db       3664 FSSGITGCIKNLVLHTARPGAPPPQPLDLQHRAQAGANTRPCPS 3707


RESULT 3
T19821
hypothetical protein ZC101.2e - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Sep-2002
```

```
C;Accession: T19821; T19819; T19820; T27490; T27488; T27489; T27487; A47648; B47648; C47
R;Baynes, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19182
A;Accession: T19821
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <WIL>
A;Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone C38C6
A;Accession: T19819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSRKTRTTTKLFGSW
A;Cross-references: EMBL:Z93375; PIDN:CAB07567.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone C38C6
A;Accession: T19820
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694,'H',1883-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPL
A;Cross-references: EMBL:Z93375; PIDN:CAB07568.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone C38C6
R;Percy, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20375
A;Accession: T27490
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <WI2>
A;Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone ZC101
A;Accession: T27488
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSRKTRTTTKLFGSW
A;Cross-references: EMBL:Z93395; PIDN:CAB07706.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone ZC101
A;Accession: T27489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694,'H',1883-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPL
A;Cross-references: EMBL:Z93395; PIDN:CAB07707.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone ZC101
A;Accession: T27487
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1128,1290,'DFARNSPS',1299,'NSS',1303-1304,'R','RHR',1544-1545,'RIRVRS',155
A;Cross-references: EMBL:Z93395; PIDN:CAB07704.1; GSPDB:GN00020; CESP:ZC101.2b
A;Experimental source: clone ZC101
R;Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
Genes Dev. 7, 1471-1484, 1993
A;Title: Products of the unc-52 gene in Caenorhabditis elegans are homologous to the cor
A;Reference number: A47648; MUID:93339574; PMID:8393416
A;Accession: A47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSR
A;Cross-references: GB:L13458
A;Accession: B47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-2198,'D',2290,'NAR',2294,'L',2296,'WHATE',2302-2303,'V',2305,'
1,'ANIV',2516-2517,'LQQG',2522,'IDG',2526,'S',2528,'SRGFHV',2535,'F' <RO2>
A;Cross-references: GB:L13458
A;Accession: C47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-1128,1290,'DFARNSPS',1299,'NSS',1303-1304,'R','RHR',1544-1545,
A;Cross-references: GB:L13458
C;Genetics:
A;Gene: CESP:ZC101.2e; CESP:ZC101.2a; CESP:ZC101.2c; CESP:ZC101.2b
A;Map position: 2
A;Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2; 1129/1; 1158/3
```

```
2; 2613/1; 2684/1; 2757/1; 2813/3; 2863/1; 2900/3; 3084/1; 3176/1; 3250/2
C;Superfamily: LR11 protein; laminin-type EGF-like homology; LDL receptor ligand-binding
C;Keywords: extracellular matrix
F;149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;955-1002/Domain: laminin-type EGF-like homology <LEG1>
F;1011-1058/Domain: laminin-type EGF-like homology <LEG2>

  Query Match             22.3%;  Score 349;  DB 2;  Length 3375;
  Best Local Similarity   31.2%;  Pred. No. 4.5e-17;
  Matches   81;  Conservative   41;  Mismatches  102;  Indels   36;  Gaps    6;

Qy          30 GDLCEHEE--NPCQLREPCLHGGTCQGTR-----CLCLPGFSGPRCQQGSGHGIAESDWH 82
               ||:   :|  | |     |   | |         | |   : || |
Db        3129 GDVYSTQEPNNIC-ANSTCGMNGQCVPRNMTHYTCECKLYYDGPTCSLFK---------- 3177

Qy          83 LEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVE 142
                        | :: | |  | |:      |     |  | :  : :|   :|:||||| :
Db        3178 ---------PIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-Q 3227

Qy         143 VGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQV 202
                    | :|:||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::
Db        3228 RPTVQQMEDYISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRI 3287

Qy         203 DGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAP 262
               |    | ||| |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :
Db        3288 DNYREVDGRSTGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK---- 3343

Qy         263 PPQPLDLQHRAQAGANTRPC 282
                   |||   |  | | :||
Db        3344 ----LDLMATAIDGKNVKPC 3359


RESULT 4
AGRT
agrin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0399; A38856
R;Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
Neuron 6, 811-823, 1991
A;Title: Structure and expression of a rat agrin.
A;Reference number: JH0399; MUID:91222570; PMID:1851019
A;Accession: JH0399
A;Molecule type: mRNA
A;Residues: 1-1779;1799-1959 <RUP>
A;Cross-references: GB:M64780; NID:g202798; PIDN:AAA40703.1; PID:g202800
A;Experimental source: embryonic spinal cord
A;Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator
R;Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
J. Neurosci. 12, 3535-3544, 1992
A;Title: Structure and chromosomal localization of the mammalian agrin gene.
A;Reference number: A38856; MUID:92407628; PMID:1326608
A;Accession: A38856
A;Molecule type: mRNA
A;Residues: 1780-1798 <RU2>
A;Cross-references: GB:S44194
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine re
C;Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.
ylcholine receptor clustering activity.
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-1959/Product: agrin, form 1 #status predicted <AG1>
F;1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
F;1-1779,1799-1959/Product: agrin, form 3 #status predicted <AG3>
F;1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
F;1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
F;22-50/Region: hydrophobic
F;88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
F;163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
F;236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
F;307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
```

```
F;381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
F;446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
F;511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
F;540-542/Region: motor neuron attachment (L-R-E) motif
F;596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
F;688-739/Domain: laminin-type EGF-like homology <LE1>
F;742-786/Domain: laminin-type EGF-like homology <LE2>
F;814-864/Domain: Kazal proteinase inhibitor homology <KPI9>
F;869-992/Region: serine/threonine-rich
F;1084-1086/Region: motor neuron attachment (L-R-E) motif
F;1147-1215/Region: serine/threonine-rich
F;1224-1257/Domain: EGF homology <EG1>
F;1287-1442/Domain: laminin G repeat homology <LG1>
F;1444-1476/Domain: EGF homology <EG2>
F;1483-1515/Domain: EGF homology <EG3>
F;1555-1706/Domain: laminin G repeat homology <LG2>
F;1713-1747/Domain: EGF homology <EG4>
F;1807-1959/Domain: laminin G repeat homology <LG3>
F;97-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454-473
-1476,1483-1494,1488-1504,1506-1515/Disulfide bonds: #status predicted
F;145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             21.3%;  Score 333.5;  DB 1;  Length 1959;
  Best Local Similarity   36.0%;  Pred. No. 3.4e-16;
  Matches   94;  Conservative   28;  Mismatches  108;  Indels   31;  Gaps   10;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
              |   ||  || |       | | |  |  |  |  |::|||   ||     |:     |
Db       1444 CLPNPCHGGALCQALEAGMFLCQCPPGRFGPTCADEKSPCQ-PNPCHGAAPCRVLSSGGA 1502

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
              :| |  | ||  ||             || :|       : | |:   :|   | | | |
Db       1503 KCECPLGRSGTFCQT-----------VLETAGSR----PFLADFNGFSYLELKGLHTFER 1547

Qy        115 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 174
               | |    :|:       |||||: | :    |:| ||:|| | : || | | || | |
Db       1548 DLGE-KMALEMVFLARGPSGLLLYNGQKTD--GKG-DFVSLALHNRHLEFCYDLGKGAAV 1603

Qy        175 LVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPG----PNVAVNAKGSVYIGGA 230
              : |::||  | | ||   | ||:|::||     | | ||     |:  :| |  :|||||
Db       1604 IRSKEPIALGTWVRVFLERNGRKGALQVGDGPRVLGESPKSRKVPHTMLNLKEPLYIGGA 1663

Qy        231 PDVATLT-GGRFSSGITGCVK 250
              || : |  |   ||| :| ::
Db       1664 PDFSKLARGAAVSSGFSGVIQ 1684

RESULT 5
T43060
agrin - electric ray (Discopyge ommata) (fragment)
C;Species: Discopyge ommata
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
C;Accession: T43060
R;Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMaha
submitted to the EMBL Data Library, September 1992
A;Reference number: Z22308
A;Accession: T43060
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1328 <SMI>
A;Cross-references: EMBL:L01423; NID:g213102; PID:g213103; PIDN:AAA49224.1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: glycoprotein; neuromuscular junction

  Query Match             20.1%;  Score 314.5;  DB 2;  Length 1328;
  Best Local Similarity   31.4%;  Pred. No. 5.4e-15;
  Matches   85;  Conservative   44;  Mismatches  105;  Indels   37;  Gaps   12;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
              |   ||::   |       | |     | |  |  : |||    ||     |      |:
Db        830 CSPNPCKNRGKCHMKEAEMFHCESVGEFSGPTCADKHNPCD-PNPCHQSANCMVLPEGGS 888

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
```

```
              :| |  |  |  |::     ::|::         |    :   |:   :|   | | |
Db        889 KCECPMGREGELCER-----VSEAE--------QDQGKAFIPEFNGLSYLEMNGIHTF-- 933

Qy        115 SLPEVPETIELEV--RTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGE 172
               : :: : : :||       :|:: : | :    |:| ||:|| |:||:| |:| || |
Db        934 -VSDLLQKLSMEVIFLAKDPNGMIFYNGQKTD--GRG-DFVSLNLRDGYLEFKYDLGKGA 989

Qy        173 ARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP----GPNVAVNAKGSVYIG 228
              | | |: ||    |: || | ||:| :::: :||||| ||     |: |:| | : |:|
Db        990 AVLRSKAPIPLNVWNVVTVERNGRKGLMKINKDELVSGESPKSRKAPHTALNLKEAFYVG 1049

Qy        229 GAPD---VATLTGGRFSSGITGCVKNLVLHS 256
              ||||    |   |    || || :: | | |
Db       1050 GAPDFNKFARAAG--IISGFTGAIQKLSLKS 1078

RESULT 6
AGCH
agrin precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0591; A38857; B38857; I50692
R;Tsim, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
Neuron 8, 677-689, 1992
A;Title: cDNA that encodes active agrin.
A;Reference number: JH0591; MUID:92232297; PMID:1314620
A;Accession: JH0591
A;Molecule type: mRNA
A;Residues: 1-1955 <TSI>
A;Cross-references: GB:M94271; NID:g211120; PIDN:AAA48585.1; PID:g211121
A;Experimental source: brain
R;Ruegg, M.A.; Tsim, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; Gensch, E.M.; McMaha
Neuron 8, 691-699, 1992
A;Title: The agrin gene codes for a family of basal lamina proteins that differ in funct
A;Reference number: A38857; MUID:92232298; PMID:1314621
A;Contents: alternative splicing
A;Accession: A38857
A;Molecule type: mRNA
A;Residues: 1132-1783;1795-1955 <RU2>
A;Cross-references: GB:M97371
A;Accession: B38857
A;Molecule type: mRNA
A;Residues: 1221-1647;1652-1783;1794-1955 <RU3>
A;Cross-references: GB:M97372
A;Note: translation of the nucleotide sequence is not complete
R;Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
Dev. Biol. 158, 523-535, 1993
A;Title: Developmental expression and alternative splicing of chick agrin RNA.
A;Reference number: I50692; MUID:93345745; PMID:8393816
A;Accession: I50692
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'SHLSNEIPA',1784-1795 <THO>
A;Cross-references: EMBL:U07271; NID:g459665; PIDN:AAA16788.1; PID:g459666
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C;Comment: Alternative splicing produces two inactive proteins: agrin-related protein 1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-38/Domain: signal sequence #status predicted <SIG>
F;39-1955/Product: agrin #status predicted <MAT>
F;39-1783,1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
F;39-1647,1652-1783,1794-1955/Product: agrin-related protein 2 #status predicted <AG2>
F;77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
F;152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
F;225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
F;295-344/Domain: Kazal proteinase inhibitor homology <KPI4>
F;370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
F;435-483/Domain: Kazal proteinase inhibitor homology <KPI6>
F;500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
F;584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
F;675-726/Domain: laminin-type EGF-like homology <LE1>
F;729-773/Domain: laminin-type EGF-like homology <LE2>
F;801-851/Domain: Kazal proteinase inhibitor homology <KPI9>
```

```
F;856-995/Region: serine/threonine-rich
F;1150-1219/Region: serine/threonine-rich
F;1233-1264/Domain: EGF homology <EG1>
F;1294-1448/Domain: laminin G repeat homology <LG1>
F;1429-1431/Region: motor neuron attachment (L-R-E) motif
F;1450-1482/Domain: EGF homology <EG2>
F;1489-1521/Domain: EGF homology <EG3>
F;1560-1711/Domain: laminin G repeat homology <LG2>
F;1718-1751/Domain: EGF homology <EG4>
F;1803-1955/Domain: laminin G repeat homology <LG3>
F;86-105,94-126,160-180,169-201,233-252,241-273,304-323,312-344,378-397,386-418,443-462,
1482,1489-1500,1494-1510,1512-1521/Disulfide bonds: #status predicted
F;390,659,764,814/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             19.3%;  Score 301.5;  DB 1;  Length 1955;
  Best Local Similarity   27.4%;  Pred. No. 7.1e-14;
  Matches   87;  Conservative   38;  Mismatches  108;  Indels   85;  Gaps   13;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTC-----QGT 55
              |   || |||:|       | | |   : |  |  | |||    ||    ||      |
Db       1450 CHPNPCHHGASCHVKEAEMFHCECLHSYTGPTCADERNPCD-PTPCHISATCLVLPEGGA 1508

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRS 115
               | |  |  |  |::     : | |                                  :
Db       1509 MCACPMGREGEFCER-----VTEQD----------------------------------HT 1530

Qy        116 LPEVPE----------------------TIELEVRTSTASGLLLWQGVEVGEAGQGKDFI 153
              :| :||                      ::|:     : ||:: : | :    |:| ||:
Db       1531 MPFLPEFNGFSYLELNGLQTLFLTCRQMSMEVVFLAKSPSGMIFYNGQKTD--GKG-DFV 1587

Qy        154 SLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 213
              || | ||:| :|| || | | | |::|:    |  |   | ||:| ::::  | | | ||
Db       1588 SLALHDGYLEYRYDLGKGAAVLRSKEPVPLNTWISVLLERSGRKGVMRINNGERVMGESP 1647

Qy        214 G----PNVAVNAKGSVYIGGAPDVATLT-GGRFSSGITGCVKNLVLHSARPGAPPPQPLD 268
                   |:  :| |   |:|||||  : |      |:   | |: : :     | |    |
Db       1648 KSRKVPHAFLNLKEPFYVGGAPDFSKLARAAAISTSFYGAVQRISI----KGVP---LLK 1700

Qy        269 LQH-RAQAGANT---RPC 282
               || |:    :|    ||
Db       1701 EQHIRSAVEISTFRAHPC 1718


RESULT 7
T42218
slit-1 protein homolog - rat
N;Alternate names: MEGF4 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002
C;Accession: T42218
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089; PMID:9693030
A;Accession: T42218
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1531 <NAK>
A;Cross-references: EMBL:AB011530; NID:g3449289; PIDN:BAA32460.1; PID:g3449290
A;Experimental source: strain Sprague-Dawley; brain
C;Genetics:
A;Gene: MEGF4
C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r

  Query Match             18.1%;  Score 283.5;  DB 2;  Length 1531;
  Best Local Similarity   27.9%;  Pred. No. 1.1e-12;
  Matches   81;  Conservative   48;  Mismatches   92;  Indels   69;  Gaps   14;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCE---HEENPCQLREPCLHGGTC--QGT 55
              |:   ||:|| |:      : ||| :|: | |||      | |:  | | :|  |  ||:
Db       1087 CKDHQCQNGAQCVDEIN-SYACLCAEGYSGQLCEIPPAPRNSCEGTE-CQNGANCVDQGS 1144

Qy         56 R--CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFS 113
```

```
              |  | ||||| || |::       : | :|:              | |
Db       1145 RPVCQCLPGFGGPECEKLLSVNFVDRDTYLQ--------------FTD------------ 1178

Qy        114 RSLPEVPE-TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGS-G 171
                |   |   | |:| |:  :|:||: |          | |:: |  ||:   |  ||
Db       1179 --LQNWPRANITLQVSTAEDNGILLYNG--------DNDHIAVELYQGHVRVSYDPGSYP 1228

Qy        172 EARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP 231
               : : | : ||||::| |  :   :  :: :||   ::  : | :  :|::  :|:|| |
Db       1229 SSAIYSAETINDGQFHTVELVTFDQMVNLSIDGGSPMTMDNFGKHYTLNSEAPLYVGGMP 1288

Qy        232 -DVAT--------LTGGRFSSGITGCVKNLVLHS---------ARPGAPP 263
               || :        | |  |     ||::|| :::          :||  |
Db       1289 VDVNSAAFRLWQILNGTSFH----GCIRNLYINNELQDFTKTQMKPGVVP 1334


RESULT 8
T20721
hypothetical protein F25F2.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 19-May-2000
C;Accession: T20721; T21343; T23842
R;Ainscough, R.
submitted to the EMBL Data Library, August 1994
A;Reference number: Z19314
A;Accession: T20721
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4307 <WIL>
A;Cross-references: EMBL:Z35662; PIDN:CAA84721.1; GSPDB:GN00021; CESP:F25F2.2
A;Experimental source: clone F10G11
R;Ainscough, R.
submitted to the EMBL Data Library, August 1994
A;Reference number: Z19410
A;Accession: T21343
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4307 <WI2>
A;Cross-references: EMBL:Z35599; PIDN:CAA84661.1; GSPDB:GN00021; CESP:F25F2.2
A;Experimental source: clone F25F2
R;Sulston, J.
submitted to the EMBL Data Library, June 1994
A;Reference number: Z19806
A;Accession: T23842
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4307 <WI3>
A;Cross-references: EMBL:Z34802; PIDN:CAA84339.1; GSPDB:GN00021; CESP:F25F2.2
A;Experimental source: clone M88
C;Genetics:
A;Gene: CESP:F25F2.2
A;Map position: 3
A;Introns: 79/3; 167/3; 251/3; 328/3; 587/3; 625/3; 737/3; 1006/1; 1190/2; 1494/2; 1770/
C;Superfamily: rat MEGF1 protein; cadherin repeat homology; EGF homology; laminin G repe
F;3761-3900/Domain: laminin G repeat homology <LGR>

  Query Match             16.9%;  Score 264;  DB 2;  Length 4307;
  Best Local Similarity   26.4%;  Pred. No. 8.6e-11;
  Matches   81;  Conservative   45;  Mismatches   99;  Indels   82;  Gaps   13;

Qy         21 QCLCRDGFKGDLCEHEENPC--------QLREPCLHGGTCQGTRCLCLPGFSGPRCQ--- 69
              |||| ||| |  || | | |        ||  | :|  | :   |:|  |  | :|
Db       3670 QCLCPDGFGGKRCEVETNQCAKSPCEQWQLCIPSVHNSTYE---CVCPLGMEGDKCSVPS 3726

Qy         70 -QGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET---IEL 125
               |  |  : |::  : | |                        |  ||    ||   :|:
Db       3727 CQNDGKCLEEAELSVGGDG------------------------YFEISLSNEIETRMELEI 3763

Qy        126 EVRTSTASGLLLWQGVEVGEAGQGK-DFISLGLQDGHLVFRYQLGSGEARLVSEDPINDG 184
              |::|:| :|:::|         :|| ||  | | :|   : :  |:|   : |:  : ||
Db       3764 ELKTTTHNGIIMW--------SRGKRDFHMLRLVNGTPEYHWNAGTGTGIVTSKTSVVDG 3815

Qy        185 EWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNA---KGSVYIGGAPDVATLTGGRF 241
```

```
              :|||:   |  ||  : || |:|    ||  :  :|       : :|     | :  |
Db       3816 QWHRIAISRRQRRTRMTVDDEDLQEAFSPIGSTVINLHRYSQKLVLG-----AKVDDGEL 3870

Qy        242 SSGITGCVK--------------NLVLHSARPG-----APPPQPLDLQHR----AQAGAN 278
              : |:: | :               : |  |:||     : |   |  ||     :|  ::
Db       3871 TDGVSACFRTISVDGMKVLKTRQGMKLFGAQPGCSALTSSPCNDLPCQHAGTCISQGKSH 3930

Qy        279 TR-PCPS 284
               :  |||
Db       3931 FKCECPS 3937

RESULT 9
IJFFTM
cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
C;Accession: A41087; B41087
R;Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.; Bryant, P.J.; Goodman, C.S.
Cell 67, 853-868, 1991
A;Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe
A;Reference number: A41087; MUID:92069752; PMID:1959133
A;Accession: A41087
A;Molecule type: mRNA
A;Residues: 143-485;1279-5147 <MAH>
A;Cross-references: GB:M80537
A;Accession: B41087
A;Molecule type: DNA
A;Residues: 1-142;487-1278 <MA2>
A;Cross-references: GB:M80537
A;Note: 1229-Gly and 1233-Ser were also found
C;Genetics:
A;Gene: fat
A;Cross-references: FlyBase:FBgn0001075
C;Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
C;Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
F;36-4583/Domain: extracellular #status predicted <EXT>
F;51-156/Domain: cadherin repeat homology <CR1>
F;159-270/Domain: cadherin repeat homology <CR2>
F;271-382/Domain: cadherin repeat homology <CR3>
F;390-494/Domain: cadherin repeat homology <CR4>
F;497-599/Domain: cadherin repeat homology <CR5>
F;602-708/Domain: cadherin repeat homology <CR6>
F;718-822/Domain: cadherin repeat homology <CR7>
F;831-942/Domain: cadherin repeat homology <CR8>
F;948-1049/Domain: cadherin repeat homology <CR9>
F;1052-1153/Domain: cadherin repeat homology <C10>
F;1156-1278/Domain: cadherin repeat homology <C11>
F;1281-1384/Domain: cadherin repeat homology <C12>
F;1387-1489/Domain: cadherin repeat homology <C13>
F;1492-1601/Domain: cadherin repeat homology <C14>
F;1607-1713/Domain: cadherin repeat homology <C15>
F;1717-1823/Domain: cadherin repeat homology <C16>
F;1826-1922/Domain: cadherin repeat homology <C17>
F;1925-2027/Domain: cadherin repeat homology <C18>
F;2028-2167/Domain: cadherin repeat homology <C19>
F;2169-2278/Domain: cadherin repeat homology <C20>
F;2281-2384/Domain: cadherin repeat homology <C99>
F;2387-2491/Domain: cadherin repeat homology <C21>
F;2494-2596/Domain: cadherin repeat homology <C22>
F;2599-2703/Domain: cadherin repeat homology <C23>
F;2707-2810/Domain: cadherin repeat homology <C24>
F;2813-2913/Domain: cadherin repeat homology <C25>
F;2915-3013/Domain: cadherin repeat homology <C26>
F;3014-3124/Domain: cadherin repeat homology <C27>
F;3127-3229/Domain: cadherin repeat homology <C28>
F;3232-3334/Domain: cadherin repeat homology <C29>
F;3337-3439/Domain: cadherin repeat homology <C30>
F;3442-3545/Domain: cadherin repeat homology <C31>
F;3548-3651/Domain: cadherin repeat homology <C32>
F;3654-3756/Domain: cadherin repeat homology <C33>
```

```
F;3954-4010/Domain: EGF homology <EG1>
F;4017-4048/Domain: EGF homology <EG2>
F;4056-4089/Domain: EGF homology <EG3>
F;4096-4127/Domain: EGF homology <EG4>
F;4584-4609/Domain: transmembrane #status predicted <TMM>
F;4610-5147/Domain: intracellular #status predicted <INT>

  Query Match             16.9%;  Score 264;  DB 1;  Length 5147;
  Best Local Similarity   29.3%;  Pred. No. 1e-10;
  Matches   93;  Conservative   39;  Mismatches  103;  Indels   82;  Gaps   19;

Qy          1 CERQPCQHGATCM--PAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTC----QG 54
              |  :||::| :|   | |   | |||| ||:|: ||   : |:   |||||| |     |
Db       4056 CYSKPCRNGGSCQRSPDGSSYF-CLCRPGFRGNQCESVSDSCR-PNPCLHGGLCVSLKPG 4113

Qy         55 TRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSR 114
               :| | ||  |  |:: |                      ||  |    :: ||
Db       4114 YKCNCTPGRYGRHCERFS----------------------YG--FQPLSYMTFP------ 4143

Qy        115 SLPEVPETIELEVRTSTASGLLLWQ-GVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEA 173
              :|      | :   |:  : |||:  |:   ::|   ||::: |      |     || |
Db       4144 ALDVTTNDISIVFATTKPNSLLLYNYGM---QSGGRSDFLAIELVHDRAYF----SSGGA 4196

Qy        174 R-----LVSEDPINDGEWHRVTALREGRRGSIQV-----DGE---ELVSGRSP------G 214
              |     :::   : || ||:||| | ||  |: |      |:   | : | |       |
Db       4197 RTAISTVIAGRNLADGGWHKVTATRNGRVMSLSVAKCADSGDVCTECLPGDSSCYADEVG 4256

Qy        215 PNVAVN-AKGSVYIGG---APDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQ 270
              |   :|  |  : |||   |  :    |   |  : ||     |||   |    : |:|
Db       4257 PVGTLNFNKQPLMIGGLSSADPILERPGQVHSDDLVGC-----LHSVHIGG---RALNLS 4308

Qy        271 HRAQ-----AGANTRPC 282
                 |     || | : |
Db       4309 SPLQQKGILAGCNRQAC 4325

RESULT 10
T22025
hypothetical protein F40E10.4 - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T22025
R;Smye, R.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19503
A;Accession: T22025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-601 <WIL>
A;Cross-references: EMBL:Z69792; PIDN:CAA93668.1; GSPDB:GN00028; CESP:F40E10.4
A;Experimental source: clone F40E10
C;Genetics:
A;Gene: CESP:F40E10.4
A;Map position: X

  Query Match             15.9%;  Score 249;  DB 2;  Length 601;
  Best Local Similarity   26.7%;  Pred. No. 1.3e-10;
  Matches   84;  Conservative   48;  Mismatches  115;  Indels   68;  Gaps   17;

Qy          5 PCQHGATCMPA-GEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTC----QGTRCLC 59
              ||::   |:|  | |   |:|  || |: ||  :  |:  | | :||:|        |||
Db        186 PCENNGKCIPINGSY--SCMCSPGFTGNNCETNIDDCKNVE-CQNGGSCVDGILSYDCLC 242

Qy         60 LPGFSG------------------PRCQQ---GSGHGIA---ESDWHL---EGSGGNDAPGQ 94
               ||::|                  |||   | |  :|    ||:     ||  |     |
Db        243 RPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCDRQ 302

Qy         95 YGAYFHDDG-FLAFPGHVFSRSLPEVPE-TIELEVRTSTASGLLLWQGVEVGEAGQGKDF 152
                  | : | :||          |   : || : :||::  |:||: | :         |
Db        303 MSVGFKNPGAYLALD--------PLASDGTITMTLRTTSKIGILLYYGDD--------HF 346

Qy        153 ISLGLQDGHLVFRYQLGSGEA-RLVSEDPINDGEWHRVTALREGRRGSIQVDGE--ELVS 209
              :|  | || :   | :|:  |  : |   :|||  ||::     |:  :|:|    ::|
```

```
Db        347 VSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVE 406

Qy        210 GRSPGPNVAVNAKGSVYIGGAP-----DVATLTGGRFSSGITGCVKNLVLHSARPGAPPP 264
                     :    |  :|||| |     |       : |  : ||: :: ::
Db        407 NSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEV------- 459

Qy        265 QPLDLQHRAQAGANT 279
               |::|| :|    ||
Db        460 -PINLQ-QALENVNT 472

RESULT 11
D89711
protein F40E10.4 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C;Accession: D89711
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: D89711
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-601 <STO>
A;Cross-references: GB:chr_X; PIDN:CAA93668.1; PID:g3877014; GSPDB:GN00028; CESP:F40E10.
C;Genetics:
A;Gene: F40E10.4
A;Map position: X

  Query Match             15.9%;  Score 249;  DB 2;  Length 601;
  Best Local Similarity   26.7%;  Pred. No. 1.3e-10;
  Matches   84;  Conservative   48;  Mismatches  115;  Indels   68;  Gaps   17;

Qy          5 PCQHGATCMPA-GEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTC----QGTRCLC 59
              ||::   |:|  | |   |:|  || |: ||   : |:  | | :||:|        |||
Db        186 PCENNGKCIPINGSY--SCMCSPGFTGNNCETNIDDCKNVE-CQNGGSCVDGILSYDCLC 242

Qy         60 LPGFSG----------------PRCQQ---GSGHGIA---ESDWHL---EGSGGNDAPGQ 94
               ||::|                  |||   | |  :|    ||:     ||  |     |
Db        243 RPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCDRQ 302

Qy         95 YGAYFHDDG-FLAFPGHVFSRSLPEVPE-TIELEVRTSTASGLLLWQGVEVGEAGQGKDF 152
                  | : | :||          |   : || : :||::  |:||: | :         |
Db        303 MSVGFKNPGAYLALD--------PLASDGTITMTLRTTSKIGILLYYGDD--------HF 346

Qy        153 ISLGLQDGHLVFRYQLGSGEA-RLVSEDPINDGEWHRVTALREGRRGSIQVDGE--ELVS 209
              :|  | || :   | :|:  |  : |   :|||  ||::     |:  :|:|    ::|
Db        347 VSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVE 406

Qy        210 GRSPGPNVAVNAKGSVYIGGAP-----DVATLTGGRFSSGITGCVKNLVLHSARPGAPPP 264
                     :    |  :|||| |     |       : |  : ||: :: ::
Db        407 NSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEV------- 459

Qy        265 QPLDLQHRAQAGANT 279
               |::|| :|    ||
Db        460 -PINLQ-QALENVNT 472

RESULT 12
T13953
MEGF5 protein - rat
N;Alternate names: slit protein homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 16-Aug-2002
C;Accession: T13953
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089; PMID:9693030
A;Accession: T13953
```

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1523 <NAK>
A;Cross-references: EMBL:AB011531; NID:g3449291; PIDN:BAA32461.1; PID:g3449292
C;Genetics:
A;Gene: MEGF5
C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r

  Query Match             15.7%;  Score 246;  DB 2;  Length 1523;
  Best Local Similarity   27.1%;  Pred. No. 5.8e-10;
  Matches   79;  Conservative   40;  Mismatches  106;  Indels   66;  Gaps   13;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEH-------EENPCQLREPCLHGGTC- 52
              |    |:||| |: |    : |:|  || |  |||       : :||   | | :|  |
Db       1078 CVAHKCRHGAQCVDAVN-GYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYE-CQNGAQCI 1135

Qy         53 ---QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG 109
                 |   | | |||:||||::         | ::| :     |
Db       1136 VVQQEPTCRCPPGFAGPRCEKLITVNFVGKDSYVELASAKVRP----------------- 1178

Qy        110 HVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRY-QL 168
                           | |:| |   :|:||::|          | ::| |  ||:   |  |
Db       1179 ----------QANISLQVATDKDNGILLYKG--------DNDPLALELYQGHVRLVYDSL 1220

Qy        169 GSGEARLVSEDPINDGEWHRVTALREGRRGSIQVD-GEELVSGR-SPGPNVAVNAKGSVY 226
               |    : | : :|||::| |  :   :  :: || |     |:    | | :|:   :|
Db       1221 SSPPTTVYSVETVNDGQFHSVELVMLNQTLNLVVDKGAPKSLGKLQKQPAVGINS--PLY 1278

Qy        227 IGGAPDVATLTG--------GRFSSGITGCVKNLVLHSARPG--APPPQPL 267
              :|| |   | ||         |   |  ||:  : :::      | ||| |
Db       1279 LGGIP---TSTGLSALRQGADRPLGGFHGCIHEVRINNELQDFKALPPQSL 1326

RESULT 13
T42626
secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment)
N;Alternate names: neurogenic extracellular slit protein
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 16-Aug-2002
C;Accession: T42626
R;Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
Mech. Dev. 79, 57-72, 1998
A;Title: Distinct but overlapping expression patterns of two vertebrate slit homologs im
A;Reference number: Z22177; MUID:99279238; PMID:10349621
A;Accession: T42626
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1025 <HOL>
A;Cross-references: EMBL:AF074960; NID:g4151258; PID:g4151259; PIDN:AAD04345.1
C;Genetics:
A;Gene: Slit2
C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r

  Query Match             15.5%;  Score 243.5;  DB 2;  Length 1025;
  Best Local Similarity   24.8%;  Pred. No. 5.8e-10;
  Matches   92;  Conservative   45;  Mismatches  103;  Indels  131;  Gaps   18;

Qy          1 CERQPCQHGATC-MPAGEYE-FQCLCRDGFKGDLCE-----HEENPC------------- 40
              |   ||:|| || :  ||   | | | |||:|: ||      |:| |
Db        457 CISNPCKHGGTCHLKEGENAGFWCTCADGFEGENCEVNIDDCEDNDCENNSTCVDGINNY 516

Qy         41 ---------------------QLREPCLHGGTC----QGTRCLCLPGFSGP--------- 66
                                   |   || |   |    :| :| | ||: |
Db        517 TCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYIGEHCDIDFDDC 576

Qy         67 ---RCQQGSGHGIAESDWHL---EGSGG-----------------------NDA------ 91
                 :|: |:    | : :     ||  |                       | |
Db        577 QDNKCKNGAHCTDAVNGYTCVCPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIIRI 636

Qy         92 --------PGQYGA---------YFHDDGFLAFPGHVFSRSLPEVPET-IELEVRTSTAS 133
                      ||  |          : : : :|  |      |    |:| | |:: |   |
Db        637 NEPICQCLPGYLGEKCEKLVSVNFVNKESYLQIP------SAKVRPQTNITLQIATDEDS 690
```

```
Qy          134 GLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV-SEDPINDGEWHRVTAL 192
                |:||::|         || |:: |  | :   |  ||  |  : | : |||| :| |  |
Db          691 GILLYKG--------DKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELL 742

Qy          193 REGRRGSIQVDG--EELVSGRSPGPNVAVNAKGSVYIGGAP---DVATL--TGGRFSSGI 245
                      |: |||   ::::  |      :|    :|:|| |   :||:|    |:  :
Db          743 TLDSSLSLSVDGGSPKVITNLS--KQSTLNFDSPLYVGGMPGKNNVASLRQAPGQNGTSF 800

Qy          246 TGCVKNLVLHS 256
                 ||::|| ::|
Db          801 HGCIRNLYINS 811

RESULT 14
MMHUMH
laminin alpha-2 chain - human (fragment)
N;Alternate names: laminin M chain; merosin heavy chain
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
C;Accession: PX0082; A35899; A38970; S14461
R;Hori, H.; Kanamori, T.; Mizuta, T.; Yamaguchi, N.; Liu, Y.; Nagai, Y.
J. Biochem. 116, 1212-1219, 1994
A;Title: Human laminin M chain: Epitope analysis of its monoclonal antibodies by immunos
A;Reference number: PX0082; MUID:95221315; PMID:7535762
A;Accession: PX0082
A;Molecule type: mRNA
A;Residues: 1-1751 <HOR>
A;Experimental source: placenta
R;Ehrig, K.; Leivo, I.; Argraves, W.S.; Ruoslahti, E.; Engvall, E.
Proc. Natl. Acad. Sci. U.S.A. 87, 3264-3268, 1990
A;Title: Merosin, a tissue-specific basement membrane protein, is a laminin-like protein
A;Reference number: A35899; MUID:90238994; PMID:2185464
A;Accession: A35899
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'V',623-1751 <EHR1>
A;Cross-references: EMBL:M59832
A;Accession: A38970
A;Molecule type: protein
A;Residues: 1368-1384;1389-1406;1593-1607 <EHR2>
A;Note: the sequence from Fig. 1 is inconsistent with that from Fig. 2 in lacking 1599-V
R;Ehrig, K.; Leivo, I.; Argraves, S.W.; Ruoslahti, E.; Engvall, E.
submitted to the EMBL Data Library, December 1990
A;Description: The tissue-specific basement membrane protein merosin is a laminin-like p
A;Reference number: S14461
A;Accession: S14461
A;Molecule type: mRNA
A;Residues: 'V',623-1264,'R',1266-1751 <LEI>
A;Cross-references: EMBL:M59832; NID:g187520; PIDN:AAA63215.1; PID:g187521
C;Comment: This protein is a prominent component of the basement membrane that mediates
C;Genetics:
A;Gene: GDB:LAMA2; LAMM
A;Cross-references: GDB:132362; OMIM:156225
A;Map position: 6q22-6q23
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h
C;Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; glycc
F;21-58/Domain: laminin-type EGF-like homology #status atypical <LE01>
F;61-107/Domain: laminin-type EGF-like homology <LE02>
F;110-165/Domain: laminin-type EGF-like homology <LE03>
F;168-212/Domain: laminin-type EGF-like homology <LE04>
F;527-567,1071-1300/Region: 3DM and 2D9 binding
F;811-972/Domain: laminin G repeat homology <LG1>
F;1005-1165/Domain: laminin G repeat homology <LG2>
F;1191-1354/Domain: laminin G repeat homology <LG3>
F;1430-1578/Domain: laminin G repeat homology <LG4>
F;1605-1751/Domain: laminin G repeat homology <LG5>
F;120,238,255,341,451,542,557,561,658,669,686,767,881,1001,1076,1119,1192,1199,1289,1509

  Query Match             14.3%;  Score 223.5;  DB 1;  Length 1751;
  Best Local Similarity   36.4%;  Pred. No. 2.9e-08;
  Matches   59;  Conservative   23;  Mismatches   67;  Indels   13;  Gaps    6;

Qy          122 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 181
                ||||||||   |||| :    :  |    ||  :: |::|   | | ||||:   :    |
Db         1429 TIELEVRTEAESGLLFYM-ARINHA----DFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 1483

Qy          182 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 238
                |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db         1484 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 1540

Qy          239 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 280
                |  :  | |||:|  || |   |   ||    |     || :
Db         1541 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 1580

RESULT 15
A48216
neurexin III-alpha secreted type 1 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
C;Accession: A48216; B48216
R;Ushkaryov, Y.A.; Suedhof, T.C.
Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
A;Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and
A;Reference number: A48216; MUID:93342001; PMID:8341647
A;Accession: A48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1438 <USH>
A;Cross-references: GB:L14851
A;Accession: B48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1368,1372-1438 <US2>
A;Cross-references: GB:L14851
C;Genetics:
A;Introns: 1372/1
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; brain; cell surface component; duplication; extracellu
F;1-27/Domain: signal sequence #status predicted <SIG>
F;202-234/Domain: EGF homology <EGF>
F;651-683/Domain: EGF homology <EGF1>

  Query Match             13.9%;  Score 217;  DB 2;  Length 1438;
  Best Local Similarity   26.4%;  Pred. No. 6.9e-08;
  Matches   67;  Conservative   35;  Mismatches  114;  Indels   38;  Gaps    8;

Qy           27 GFKGDLCE------------------HEENPCQLREPCLHGGTC----QGTRCLC-LPGF 63
                |||| : :                    | ||  | || :|| |        | |   |:
Db          171 GFKGLMLDLKYGNSEPRLLGSQSVQLEAEGPCGER-PCENGGICFLLDGHPTCDCSTTGY 229

Qy           64 SGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETI 123
                 |  | :    |   |   :   | : |  :  | |    :| :   :    :    : |
Db          230 GGTLCSEDVSQGPGLSHLMMSEQGRSKAREENVATFRGSEYLCY--DLSQNPIQSSSDEI 287

Qy          124 ELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI-- 181
                 |  :|   :||:|         |:  |:::| |:|| :     ||||    : | |:
Db          288 TLSFKTWQRNGLIL-------HTGKSADYVNLALKDGAVSLVINLGSGAFEAIVE-PVNG 339

Qy          182 --NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 239
                  ||  || |   |  |: :| |||    :| :      : :    |:||:|  | | |
Db          340 KFNDNAWHDVKVTRNLRQVTISVDGILTTTGYTQEDYTMLGSDDFFYVGGSPSTADLPGS 399

Qy          240 RFSSGITGCVKNLV 253
                  |:   ||:| :|
Db          400 PVSNNFMGCLKEVV 413

Search completed: March  9, 2004, 17:22:35
Job time : 10.7114 secs
```

```
                     GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.


OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:19:08 ; Search time 20.844 Seconds
                                            (without alignments)
                                            2876.963 Million cell updates/sec

Title:           US-10-006-011A-9
Perfect score:   1566
Sequence:        1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters:      809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
              1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
              2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
              3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
              4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
              5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
              6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
              7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
              8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
              9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
              10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
              11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
              12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
              13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
              14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
              15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
              16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
              17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
              18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                    %
Result             Query
  No.   Score      Match Length DB  ID                        Description
------------------------------------------------------------------------------
    1   357.5      22.8   1931  15  US-10-094-886-180          Sequence 180, App
    2     343      21.9    238   9  US-09-764-853-758         Sequence 758, App
    3     343      21.9    238   9  US-09-764-898-252         Sequence 252, App
    4     343      21.9    238  10  US-09-764-881-158          Sequence 158, App
    5     343      21.9    238  14  US-10-073-865-121          Sequence 121, App
    6     343      21.9    238  15  US-10-242-747-158          Sequence 158, App
    7     343      21.9    375   9  US-09-978-249-10          Sequence 10, Appl
    8     343      21.9    432   9  US-09-764-853-541         Sequence 541, App
    9     343      21.9    432   9  US-09-764-898-179         Sequence 179, App
   10     343      21.9    432  10  US-09-764-881-102          Sequence 102, App
   11     343      21.9    432  14  US-10-073-865-78           Sequence 78, Appl
   12     343      21.9    432  15  US-10-242-747-102          Sequence 102, App
   13     343      21.9    463  15  US-10-104-047-3058         Sequence 3058, Ap
   14     343      21.9    775  15  US-10-108-260A-4433        Sequence 4433, Ap
   15   340.5      21.7    406  15  US-10-264-049-2933         Sequence 2933, Ap
   16   333.5      21.3   1940  13  US-10-016-283-34           Sequence 34, Appl
   17     330      21.1    294  13  US-10-016-283-29           Sequence 29, Appl
   18     330      21.1    338  13  US-10-016-283-28           Sequence 28, Appl
   19     330      21.1    390  13  US-10-016-283-27           Sequence 27, Appl
   20     330      21.1    440  13  US-10-016-283-26           Sequence 26, Appl
   21     330      21.1    456  13  US-10-016-283-25           Sequence 25, Appl
   22     330      21.1    492  13  US-10-016-283-36           Sequence 36, Appl
   23     330      21.1    819  10  US-09-866-050A-503         Sequence 503, App
   24     326      20.8    256  13  US-10-016-283-30           Sequence 30, Appl
   25     292      18.6   1534  11  US-09-970-944-30           Sequence 30, Appl
   26   290.5      18.6    210   9  US-09-764-870-432         Sequence 432, App
   27   290.5      18.6    210  14  US-10-125-540-432          Sequence 432, App
   28   283.5      18.1   1531  11  US-09-970-944-31           Sequence 31, Appl
   29   280.5      17.9   1531  11  US-09-970-944-28           Sequence 28, Appl
   30   278.5      17.8    152  15  US-10-104-047-3338         Sequence 3338, Ap
   31     269      17.2   5147  14  US-10-174-677-4            Sequence 4, Appli
   32     264      16.9   4307  15  US-10-369-493-5698         Sequence 5698, Ap
   33     264      16.9   4307  15  US-10-369-493-5699         Sequence 5699, Ap
   34     264      16.9   4307  15  US-10-369-493-5700         Sequence 5700, Ap
   35     261      16.7    216  13  US-10-016-283-31           Sequence 31, Appl
   36   253.5      16.2   1525  14  US-10-289-776-2            Sequence 2, Appli
   37   253.5      16.2   1529  10  US-09-766-511B-29          Sequence 29, Appl
   38   253.5      16.2   1529  14  US-10-189-123-67           Sequence 67, Appl
   39   253.5      16.2   1529  14  US-10-188-495-67           Sequence 67, Appl
   40     252      16.1    194   9  US-09-764-870-537         Sequence 537, App
   41     252      16.1    194  14  US-10-125-540-537          Sequence 537, App
   42     249      15.9    601  15  US-10-369-493-6999         Sequence 6999, Ap
   43     249      15.9    735  14  US-10-289-776-9            Sequence 9, Appli
   44   245.5      15.7   1529  10  US-09-866-050A-396         Sequence 396, App
   45   244.5      15.6   1512  11  US-09-970-944-32           Sequence 32, Appl


                                 ALIGNMENTS


RESULT 1
US-10-094-886-180
; Sequence 180, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
;  APPLICANT: Kekuda, Ramesh
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Vernet, Corine A.
;  APPLICANT:  Li, Li
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Malyankar, Uriel M.
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Taupier, Raymond J., Jr.
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Pena, Carol
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Gusev, Vladimir
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Gerlach, Valerie
;  APPLICANT:  Pochart, Pascal
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Spaderna, Steven
;  APPLICANT:  LaRochelle, William
;  APPLICANT:  Zhong, Mei
;  TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
;  FILE REFERENCE: 21402-290 B
;  CURRENT APPLICATION NUMBER: US/10/094,886
```

```
;   CURRENT FILING DATE:  2002-03-07
;   PRIOR APPLICATION NUMBER: 60/274,322
;   PRIOR FILING DATE: 2001-03-08
;   PRIOR APPLICATION NUMBER: 60/313,182
;   PRIOR FILING DATE: 2001-08-17
;   PRIOR APPLICATION NUMBER: 60/288,052
;   PRIOR FILING DATE: 2001-05-02
;   PRIOR APPLICATION NUMBER: 60/318,510
;   PRIOR FILING DATE: 2001-09-10
;   PRIOR APPLICATION NUMBER: 60/274,281
;   PRIOR FILING DATE: 2001-03-08
;   PRIOR APPLICATION NUMBER: 60/314,018
;   PRIOR FILING DATE: 2001-08-21
;   PRIOR APPLICATION NUMBER: 60/274,194
;   PRIOR FILING DATE: 2001-03-08
;   PRIOR APPLICATION NUMBER: 60/274,849
;   PRIOR FILING DATE: 2001-03-09
;   PRIOR APPLICATION NUMBER: 60/296,693
;   PRIOR FILING DATE: 2001-06-07
;   PRIOR APPLICATION NUMBER: 60/313,626
;   PRIOR FILING DATE: 2001-08-21
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 298
;   SOFTWARE: PatentIn 2.1
; SEQ ID NO 180
;   LENGTH: 1931
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-886-180

  Query Match             22.8%;  Score 357.5;  DB 15;  Length 1931;
  Best Local Similarity   35.4%;  Pred. No. 2.3e-22;
  Matches   91;  Conservative   29;  Mismatches  110;  Indels   27;  Gaps    8;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
              |   ||  || |       | | |  |  |  |  |::|||   ||     |:     |
Db       1439 CLPNPCHGGAPCQNLEAGRFHCQCPPGRVGPTCADEKSPCQ-PNPCHGAAPCRVLPEGGA 1497

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
              :| |  |  |  ||  |               | |  | : | |:    |   | | |:|
Db       1498 QCECPLGREGTFCQTAS---------------GQDGSGPFLADFNGFSHLELRGLHTFAR 1542

Qy        115 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 174
               | |    :|:       |||||: | :    |:| ||:|| |:|  | ||| || | |
Db       1543 DLGE-KMALEVVFLARGPSGLLLYNGQKTD--GKG-DFVSLALRDRRLEFRYDLGKGAAV 1598

Qy        175 LVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVA 234
              : | :|:  | | ||:  | ||:|:::|     | | || |:  :| |  :|:||||| :
Db       1599 IRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFS 1658

Qy        235 TLT-GGRFSSGITGCVK 250
               |      |||  | ::
Db       1659 KLARAAAVSSGFDGAIQ 1675

RESULT 2
US-09-764-853-758
; Sequence 758, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PJZ06
;   CURRENT APPLICATION NUMBER: US/09/764,853
;   CURRENT FILING DATE:  2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 939
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 758
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-758

  Query Match             21.9%;  Score 343;  DB 9;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.1e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        219 HISLVEDAVDGKNINTC 235

RESULT 3
US-09-764-898-252
; Sequence 252, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PJZ01
;   CURRENT APPLICATION NUMBER: US/09/764,898
;   CURRENT FILING DATE:  2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 311
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 252
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-898-252

  Query Match             21.9%;  Score 343;  DB 9;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.1e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218
```

```
Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        219 HISLVEDAVDGKNINTC 235

RESULT 4
US-09-764-881-158
; Sequence 158, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07
;  CURRENT APPLICATION NUMBER: US/09/764,881
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-158

  Query Match             21.9%;  Score 343;  DB 10;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.1e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |:  ||: :  | :
Db        166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        219 HISLVEDAVDGKNINTC 235

RESULT 5
US-10-073-865-121
; Sequence 121, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ09C1
;  CURRENT APPLICATION NUMBER: US/10/073,865
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 154
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)
```

```
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-073-865-121

  Query Match             21.9%;  Score 343;  DB 14;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.1e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |:  ||: :  | :
Db        166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        219 HISLVEDAVDGKNINTC 235

RESULT 6
US-10-242-747-158
; Sequence 158, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07C1
;  CURRENT APPLICATION NUMBER: US/10/242,747
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: 09/764,881
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/214,886
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: 60/217,487
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,758
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/220,963
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/217,496
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,447
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/218,290
;  PRIOR FILING DATE: 2000-07-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
;   LENGTH: 238
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-242-747-158

  Query Match             21.9%;  Score 343;  DB 15;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.1e-22;
```

```
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy      34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
           |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db      11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy      90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
           : |      |  | : |      |   |||       : :  :|:   |||||:|
Db      57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy     146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db     106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy     206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
              :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db     166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy     266 PLDLQHRAQAGANTRPC 282
            : |   |  | |   |
Db     219 HISLVEDAVDGKNINTC 235


RESULT 7
US-09-978-249-10
; Sequence 10, Application US/09978249
; Patent No. US20020106780A1
; GENERAL INFORMATION:
;  APPLICANT: Fiscella, et al.
;  TITLE OF INVENTION: Extracellular Matrix Polynucleotides, Polypeptides, and Antibodie
;  FILE REFERENCE: PT054P1
;  CURRENT APPLICATION NUMBER: US/09/978,249
;  CURRENT FILING DATE:  2001-10-17
;  PRIOR APPLICATION NUMBER: PCT/US01/11643
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: 60/198,123
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 375
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-978-249-10

  Query Match             21.9%;  Score 343;  DB 9;  Length 375;
  Best Local Similarity   33.5%;  Pred. No. 5.5e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy      34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
           |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db     148 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 193

Qy      90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
           : |      |  | : |      |   |||       : :  :|:   |||||:|
Db     194 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 242

Qy     146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db     243 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 302

Qy     206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
              :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db     303 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 355

Qy     266 PLDLQHRAQAGANTRPC 282
            : |   |  | |   |
Db     356 HISLVEDAVDGKNINTC 372


RESULT 8
US-09-764-853-541
; Sequence 541, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ06
;  CURRENT APPLICATION NUMBER: US/09/764,853
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 939
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 541
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-853-541

  Query Match             21.9%;  Score 343;  DB 9;  Length 432;
  Best Local Similarity   33.5%;  Pred. No. 6.6e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy      34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
           |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db     205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250

Qy      90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
           : |      |  | : |      |   |||       : :  :|:   |||||:|
Db     251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy     146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db     300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy     206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
              :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db     360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy     266 PLDLQHRAQAGANTRPC 282
            : |   |  | |   |
Db     413 HISLVEDAVDGKNINTC 429


RESULT 9
US-09-764-898-179
; Sequence 179, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ01
;  CURRENT APPLICATION NUMBER: US/09/764,898
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 311
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 179
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-898-179

  Query Match             21.9%;  Score 343;  DB 9;  Length 432;
  Best Local Similarity   33.5%;  Pred. No. 6.6e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy      34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
           |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db     205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250

Qy      90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
           : |      |  | : |      |   |||       : :  :|:   |||||:|
Db     251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy     146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
```

```
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        413 HISLVEDAVDGKNINTC 429

RESULT 10
US-09-764-881-102
; Sequence 102, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07
;  CURRENT APPLICATION NUMBER: US/09/764,881
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-881-102

  Query Match             21.9%;  Score 343;  DB 10;  Length 432;
  Best Local Similarity   33.5%;  Pred. No. 6.6e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db        205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
Db        251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        413 HISLVEDAVDGKNINTC 429

RESULT 11
US-10-073-865-78
; Sequence 78, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ09C1
;  CURRENT APPLICATION NUMBER: US/10/073,865
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 154
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-073-865-78

  Query Match             21.9%;  Score 343;  DB 14;  Length 432;
  Best Local Similarity   33.5%;  Pred. No. 6.6e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db        205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
Db        251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        413 HISLVEDAVDGKNINTC 429

RESULT 12
US-10-242-747-102
; Sequence 102, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07C1
;  CURRENT APPLICATION NUMBER: US/10/242,747
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: 09/764,881
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/214,886
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: 60/217,487
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,758
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/220,963
;  PRIOR FILING DATE: 2000-07-26
;  PRIOR APPLICATION NUMBER: 60/217,496
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,447
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/218,290
;  PRIOR FILING DATE: 2000-07-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-242-747-102

  Query Match             21.9%;  Score 343;  DB 15;  Length 432;
  Best Local Similarity   33.5%;  Pred. No. 6.6e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db        205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250
```

```
Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         413 HISLVEDAVDGKNINTC 429

RESULT 13
US-10-104-047-3058
; Sequence 3058, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3058
;   LENGTH: 463
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3058

  Query Match             21.9%;  Score 343;  DB 15;  Length 463;
  Best Local Similarity   33.5%;  Pred. No. 7.2e-22;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         236 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 281

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         282 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 330

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         331 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 390

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         391 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 443

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         444 HISLVEDAVDGKNINTC 460

RESULT 14
US-10-108-260A-4433
; Sequence 4433, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4433
;   LENGTH: 775
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-4433

  Query Match             21.9%;  Score 343;  DB 15;  Length 775;
  Best Local Similarity   33.5%;  Pred. No. 1.4e-21;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         548 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 593

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         594 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 642

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         643 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 702

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         703 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 755

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         756 HISLVEDAVDGKNINTC 772

RESULT 15
US-10-264-049-2933
; Sequence 2933, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA133P1
;  CURRENT APPLICATION NUMBER: US/10/264,049
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/18569
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: US 60/209,467
;  PRIOR FILING DATE: 2000-06-07
;  NUMBER OF SEQ ID NOS: 4360
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2933
;   LENGTH: 406
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (82)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (95)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (96)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2933

  Query Match             21.7%;  Score 340.5;  DB 15;  Length 406;
  Best Local Similarity   29.2%;  Pred. No. 1e-21;
  Matches   99;  Conservative   38;  Mismatches  113;  Indels   89;  Gaps   11;

Qy           1 CERQPCQHGATCMP------------AGEYEFQCLCR-----DGFKG------------- 30
               | |:|   | : :|             |  :|  | |      || |
```

```
Db        101 CGRRPRVLGESPVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLL 160

Qy         31 ---------DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHG 75
                       |:     :||      |||:| :|        |||  ||||| |::|
Db        161 TPEHVLRQVDVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL--- 217

Qy         76 IAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEV 127
                     :| | |            |   ||| |  |   |  | |:         || :
Db        218 -------VEKSAG------------DVDTLAFDGRTFVEYLNAVTESEKALQSNHFELSL 258

Qy        128 RTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWH 187
              ||    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   |
Db        259 RTEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWL 314

Qy        188 RVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGI 245
              || | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db        315 RVVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGF 374

Qy        246 TGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
               ||::::|:           || |   |      ||||:
Db        375 VGCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 405
```

Search completed: March 9, 2004, 17:25:15
Job time : 21.844 secs

```
                         GenCore version 5.1.6
                Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:16:23 ; Search time 11.1326 Seconds
                                           (without alignments)
                                           1317.011 Million cell updates/sec

Title:          US-10-006-011A-9
Perfect score:  1566
Sequence:       1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                     Description
------------------------------------------------------------------------
    1   333.5    21.3   1940  2  US-08-644-271-30          Sequence 30, Appl
    2   333.5    21.3   1940  4  US-09-077-955-34          Sequence 34, Appl
    3     330    21.1    294  4  US-09-077-955-29          Sequence 29, Appl
    4     330    21.1    338  4  US-09-077-955-28          Sequence 28, Appl
    5     330    21.1    390  4  US-09-077-955-27          Sequence 27, Appl
    6     330    21.1    440  4  US-09-077-955-26          Sequence 26, Appl
    7     330    21.1    456  4  US-09-077-955-25          Sequence 25, Appl
    8     330    21.1    492  2  US-08-644-271-32          Sequence 32, Appl
    9     330    21.1    492  4  US-09-077-955-36          Sequence 36, Appl
   10     326    20.8    256  4  US-09-077-955-30          Sequence 30, Appl
   11     261    16.7    216  4  US-09-077-955-31          Sequence 31, Appl
   12   253.5    16.2   1525  3  US-09-191-647-2           Sequence 2, Appli
   13   253.5    16.2   1525  3  US-09-540-245A-2          Sequence 2, Appli
   14   253.5    16.2   1525  3  US-09-540-153-2           Sequence 2, Appli
   15     249    15.9    735  3  US-09-191-647-9           Sequence 9, Appli
   16     249    15.9    735  3  US-09-540-245A-9          Sequence 9, Appli
   17     249    15.9    735  3  US-09-540-153-9           Sequence 9, Appli
   18   245.5    15.7   1529  4  US-09-312-283C-396        Sequence 396, App
   19     239    15.3   1523  4  US-09-182-024A-2          Sequence 2, Appli
   20   226.5    14.5    716  4  US-09-312-283C-183        Sequence 183, App
   21   226.5    14.5    771  3  US-09-188-930-183         Sequence 183, App
   22   223.5    14.3   1130  2  US-08-460-309-2           Sequence 2, Appli
   23   223.5    14.3   1130  2  US-08-125-077-2           Sequence 2, Appli
   24   223.5    14.3   1130  6  5444158-2                Patent No. 5444158
   25   223.5    14.3   3088  4  US-09-562-702A-8          Sequence 8, Appli
   26   223.5    14.3   3089  4  US-09-562-702A-4          Sequence 4, Appli
   27   223.5    14.3   3110  4  US-09-562-702A-2          Sequence 2, Appli
   28   223.5    14.3   3110  4  US-09-562-702A-6          Sequence 6, Appli
   29   223.5    14.3   3110  4  US-09-561-709B-7          Sequence 7, Appli
   30   223.5    14.3   3111  2  US-08-460-309-4           Sequence 4, Appli
   31   223.5    14.3   3111  2  US-08-125-077-4           Sequence 4, Appli
   32   207.5    13.3   3084  4  US-09-562-702A-12         Sequence 12, Appl
   33   207.5    13.3   3106  4  US-09-562-702A-10         Sequence 10, Appl
   34     206    13.2   3571  4  US-09-911-842A-2          Sequence 2, Appli
   35   202.5    12.9   3594  4  US-09-911-842A-4          Sequence 4, Appli
   36   199.5    12.7    160  3  US-09-191-647-5           Sequence 5, Appli
   37   199.5    12.7    160  3  US-09-540-245A-5          Sequence 5, Appli
   38   199.5    12.7    160  3  US-09-540-153-5           Sequence 5, Appli
   39     199    12.7   2321  4  US-09-230-652-2           Sequence 2, Appli
   40   188.5    12.0   3075  2  US-08-460-309-5           Sequence 5, Appli
   41   188.5    12.0   3075  2  US-08-125-077-5           Sequence 5, Appli
   42     183    11.7   1480  3  US-09-191-647-7           Sequence 7, Appli
   43     183    11.7   1480  3  US-09-540-245A-7          Sequence 7, Appli
   44     183    11.7   1480  3  US-09-540-153-7           Sequence 7, Appli
   45     183    11.7   1480  5  PCT-US91-09055-2          Sequence 2, Appli

                                ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1940 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Rat Agrin
;      LOCATION:  1...1940
;      OTHER INFORMATION:
US-08-644-271-30

  Query Match             21.3%;  Score 333.5;  DB 2;  Length 1940;
```

```
  Best Local Similarity   36.0%;  Pred. No. 9.2e-22;
  Matches   94;  Conservative   28;  Mismatches  108;  Indels   31;  Gaps   10;

Qy            1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
                |   ||  || |       | | |  |  |  |  |::|||   ||     |:     |
Db         1444 CLPNPCHGGALCQALEAGMFLCQCPPGRFGPTCADEKSPCQ-PNPCHGAAPCRVLSSGGA 1502

Qy           56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
                :| |  | ||  ||             || :|       : | |:   :|   | | | |
Db         1503 KCECPLGRSGTFCQT-----------VLETAGSR----PFLADFNGFSYLELKGLHTFER 1547

Qy          115 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 174
                 | |    :|:       |||||: | :    |:| ||:|| | : || | | || | |
Db         1548 DLGE-KMALEMVFLARGPSGLLLYNGQKTD--GKG-DFVSLALHNRHLEFCYDLGKGAAV 1603

Qy          175 LVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPG----PNVAVNAKGSVYIGGA 230
                : |::||  | | ||   | ||:|::||     | | ||     |:  :| |  :|||||
Db         1604 IRSKEPIALGTWVRVFLERNGRKGALQVGDGPRVLGESPKSRKVPHTMLNLKEPLYIGGA 1663

Qy          231 PDVATLT-GGRFSSGITGCVK 250
                || : |  |   ||| :| ::
Db         1664 PDFSKLARGAAVSSGFSGVIQ 1684

RESULT 2
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-077-955-34

  Query Match             21.3%;  Score 333.5;  DB 4;  Length 1940;
  Best Local Similarity   36.0%;  Pred. No. 9.2e-22;
  Matches   94;  Conservative   28;  Mismatches  108;  Indels   31;  Gaps   10;

Qy            1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
                |   ||  || |       | | |  |  |  |  |::|||   ||     |:     |
Db         1444 CLPNPCHGGALCQALEAGMFLCQCPPGRFGPTCADEKSPCQ-PNPCHGAAPCRVLSSGGA 1502

Qy           56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
                :| |  | ||  ||             || :|       : | |:   :|   | | | |
Db         1503 KCECPLGRSGTFCQT-----------VLETAGSR----PFLADFNGFSYLELKGLHTFER 1547

Qy          115 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 174
                 | |    :|:       |||||: | :    |:| ||:|| | : || | | || | |
Db         1548 DLGE-KMALEMVFLARGPSGLLLYNGQKTD--GKG-DFVSLALHNRHLEFCYDLGKGAAV 1603

Qy          175 LVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPG----PNVAVNAKGSVYIGGA 230
                : |::||  | | ||   | ||:|::||     | | ||     |:  :| |  :|||||
Db         1604 IRSKEPIALGTWVRVFLERNGRKGALQVGDGPRVLGESPKSRKVPHTMLNLKEPLYIGGA 1663

Qy          231 PDVATLT-GGRFSSGITGCVK 250
                || : |  |   ||| :| ::
Db         1664 PDFSKLARGAAVSSGFSGVIQ 1684
```

```
RESULT 3
US-09-077-955-29
; Sequence 29, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 294
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-29

  Query Match             21.1%;  Score 330;  DB 4;  Length 294;
  Best Local Similarity   31.3%;  Pred. No. 1.5e-22;
  Matches   87;  Conservative   34;  Mismatches   99;  Indels   58;  Gaps    8;

Qy           31 DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLE 84
                |:     :||      |||:| :|        |||  ||||| |::|           :|
Db           50 DVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VE 99

Qy           85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVR 128
                 | |            |   ||| |  |   |  | |:                 || :|
Db          100 KSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLR 147

Qy          129 TSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHR 188
                |    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | |
Db          148 TEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLR 203

Qy          189 VTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
                | | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db          204 VVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFV 263

Qy          247 GCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
                ||::::|:           || |   |      ||||:
Db          264 GCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 293

RESULT 4
US-09-077-955-28
; Sequence 28, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-28
```

```
  Query Match             21.1%;  Score 330;  DB 4;  Length 338;
  Best Local Similarity   31.3%;  Pred. No. 1.9e-22;
  Matches   87;  Conservative   34;  Mismatches   99;  Indels   58;  Gaps    8;

Qy          31 DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||      |||:| :|        |||  ||||| |::|           :|
Db          94 DVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VE 143

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVR 128
                | |            |   ||| |  |   |  | |:                 || :|
Db         144 KSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLR 191

Qy         129 TSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHR 188
               |    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | |
Db         192 TEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLR 247

Qy         189 VTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
               | | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db         248 VVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFV 307

Qy         247 GCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
               ||::::|:           || |   |      ||||:
Db         308 GCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 337

RESULT 5
US-09-077-955-27
; Sequence 27, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
;   LENGTH: 390
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-27

  Query Match             21.1%;  Score 330;  DB 4;  Length 390;
  Best Local Similarity   31.3%;  Pred. No. 2.3e-22;
  Matches   87;  Conservative   34;  Mismatches   99;  Indels   58;  Gaps    8;

Qy          31 DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||      |||:| :|        |||  ||||| |::|           :|
Db         146 DVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VE 195

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVR 128
                | |            |   ||| |  |   |  | |:                 || :|
Db         196 KSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLR 243

Qy         129 TSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHR 188
               |    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | |
Db         244 TEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLR 299

Qy         189 VTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
               | | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db         300 VVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFV 359

Qy         247 GCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
               ||::::|:           || |   |      ||||:
Db         360 GCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 389
```

```
RESULT 6
US-09-077-955-26
; Sequence 26, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 440
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-26

  Query Match             21.1%;  Score 330;  DB 4;  Length 440;
  Best Local Similarity   31.3%;  Pred. No. 2.7e-22;
  Matches   87;  Conservative   34;  Mismatches   99;  Indels   58;  Gaps    8;

Qy          31 DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||      |||:| :|        |||  ||||| |::|           :|
Db         196 DVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VE 245

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVR 128
                | |            |   ||| |  |   |  | |:                 || :|
Db         246 KSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLR 293

Qy         129 TSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHR 188
               |    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | |
Db         294 TEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLR 349

Qy         189 VTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
               | | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db         350 VVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFV 409

Qy         247 GCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
               ||::::|:           || |   |      ||||:
Db         410 GCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 439

RESULT 7
US-09-077-955-25
; Sequence 25, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-25
```

```
  Query Match             21.1%;  Score 330;  DB 4;  Length 456;
  Best Local Similarity   31.3%;  Pred. No. 2.8e-22;
  Matches   87;  Conservative   34;  Mismatches   99;  Indels   58;  Gaps    8;

Qy          31 DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||      |||:| :|        |||  ||||| |::|           :|
Db         212 DVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VE 261

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVR 128
                | |            |   ||| |  |   |  | |:                 || :|
Db         262 KSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLR 309

Qy         129 TSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHR 188
               |    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | |
Db         310 TEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLR 365

Qy         189 VTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
               | | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db         366 VVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFV 425

Qy         247 GCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
               ||::::|:           || |   |      ||||:
Db         426 GCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 455

RESULT 8
US-08-644-271-32
; Sequence 32, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  492 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
;    FEATURE:
;      NAME/KEY:  Human Agrin
;      LOCATION:  1...492
;      OTHER INFORMATION:
US-08-644-271-32

  Query Match             21.1%;  Score 330;  DB 2;  Length 492;
  Best Local Similarity   31.3%;  Pred. No. 3.1e-22;
  Matches   87;  Conservative   34;  Mismatches   99;  Indels   58;  Gaps    8;

Qy          31 DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||      |||:| :|        |||  ||||| |::|           :|
Db         248 DVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VE 297

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVR 128
                | |            |   ||| |  |   |  | |:                 || :|
Db         298 KSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLR 345

Qy         129 TSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHR 188
               |    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | |
Db         346 TEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLR 401

Qy         189 VTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
               | | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db         402 VVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFV 461

Qy         247 GCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
               ||::::|:           || |   |      ||||:
Db         462 GCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 491

RESULT 9
US-09-077-955-36
; Sequence 36, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;   LENGTH: 492
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-36

  Query Match             21.1%;  Score 330;  DB 4;  Length 492;
  Best Local Similarity   31.3%;  Pred. No. 3.1e-22;
  Matches   87;  Conservative   34;  Mismatches   99;  Indels   58;  Gaps    8;

Qy          31 DLCEHEENPCQLR--EPCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLE 84
               |:     :||      |||:| :|        |||  ||||| |::|           :|
Db         248 DVTSFAGHPCTRASGHPCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VE 297

Qy          85 GSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVR 128
                | |            |   ||| |  |   |  | |:                 || :|
Db         298 KSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLR 345

Qy         129 TSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHR 188
               |    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | |
Db         346 TEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLR 401

Qy         189 VTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGIT 246
               | | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db         402 VVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFV 461

Qy         247 GCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
```

```
              ||::::|:           || |   |      ||||:
Db         462 GCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 491

RESULT 10
US-09-077-955-30
; Sequence 30, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 256
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-30

  Query Match             20.8%;  Score 326;  DB 4;  Length 256;
  Best Local Similarity   32.1%;  Pred. No. 3e-22;
  Matches   84;  Conservative   32;  Mismatches   90;  Indels   56;  Gaps    7;

Qy          45 PCLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 100
              |||:| :|        |||  ||||| |::|           :| | |
Db          28 PCLNGASCVPREAAYVCLCPGGFSGPHCEKGL----------VEKSAG------------ 65

Qy         101 DDGFLAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVG 144
              |   ||| |  |   |  | |:                 || :||    ||:||     |
Db          66 DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----G 121

Qy         145 EAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDG 204
              :| :  |:::| : ||||   | |||    | |  |:|   | || | || | ||:||
Db         122 KATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGN 181

Qy         205 EELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAP 262
              |  |:| ||     ::  |::::|| |:  |       : :|  ||::::|:
Db         182 EAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR------ 235

Qy         263 PPQPLDLQHRAQAGANTRPCPS 284
                 || |   |      ||||:
Db         236 --HPLHLLEDAVTKPELRPCPT 255

RESULT 11
US-09-077-955-31
; Sequence 31, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 216
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-31

  Query Match             16.7%;  Score 261;  DB 4;  Length 216;
  Best Local Similarity   32.8%;  Pred. No. 2.4e-16;
  Matches   65;  Conservative   27;  Mismatches   76;  Indels   30;  Gaps    4;

Qy         105 LAFPGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQ 148
              ||| |  |   |  | |:                 || :||    ||:||     |:| :
Db          30 LAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATE 85

Qy         149 GKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELV 208
                |:::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |
Db          86 RADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPV 145

Qy         209 SGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQP 266
              :| ||     ::  |::::|| |:  |       : :|  ||::::|:           |
Db         146 TGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HP 197

Qy         267 LDLQHRAQAGANTRPCPS 284
              | |   |      ||||:
Db         198 LHLLEDAVTKPELRPCPT 215

RESULT 12
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
;  APPLICANT: Goodman, Corey
;  APPLICANT: Kid, Thomas
;  APPLICANT: Brose, Katja
;  APPLICANT: Tessier-Lavigne, Marc
;  TITLE OF INVENTION: Modulating Robo: Ligand Interactions
;  FILE REFERENCE: B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/191,647
;  CURRENT FILING DATE: 1998-11-13
;  EARLIER APPLICATION NUMBER: 60/065,544
;  EARLIER FILING DATE: 1997-11-14
;  EARLIER APPLICATION NUMBER: 60/081,057
;  EARLIER FILING DATE: 1998-04-07
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-191-647-2

  Query Match             16.2%;  Score 253.5;  DB 3;  Length 1525;
  Best Local Similarity   24.8%;  Pred. No. 1.6e-14;
  Matches   92;  Conservative   48;  Mismatches  100;  Indels  131;  Gaps   18;

Qy           1 CERQPCQHGATC-MPAGEYE-FQCLCRDGFKGDLCE-----HEENPC------------- 40
              |   ||:|| || :  || : | |:| |||:|: ||      |:| |
Db         957 CISNPCKHGGTCHLKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNY 1016

Qy          41 ---------------------QLREPCLHGGTC----QGTRCLCLPGFSGP--------- 66
                                   |   || |   |    :| :| | ||: |
Db        1017 TCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDC 1076

Qy          67 ---RCQQGSGHGIAESDWHL---EGSGG----------------------NDA------ 91
                 :|: |:    | : :     ||  |                      | |
Db        1077 QDNKCKNGAHCTDAVNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRI 1136

Qy          92 --------PGQYGA---------YFHDDGFLAFPGHVFSRSLPEVPET-IELEVRTSTAS 133
                      ||  |          : : : :|  |      |    |:| | |:: |   |
Db        1137 NEPICQCLPGYQGEKCEKLVSVNFINKESYLQIP------SAKVRPQTNITLQIATDEDS 1190

Qy         134 GLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV-SEDPINDGEWHRVTAL 192
              |:||::|         || |:: |  | :   |  ||  |  : | : |||| :| |  |
Db        1191 GILLYKG--------DKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELL 1242
```

```
Qy        193 REGRRGSIQVDG--EELVSGRSPGPNVAVNAKGSVYIGGAP---DVATL--TGGRFSSGI 245
                 :  |: |||   ::::  |      :|    :|:|| |   :||:|    |:  :
Db       1243 ALDQSLSLSVDGGNPKIITNLS--KQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSF 1300

Qy        246 TGCVKNLVLHS 256
               ||::|| ::|
Db       1301 HGCIRNLYINS 1311

RESULT 13
US-09-540-245A-2
; Sequence 2, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,245A
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  60/065,544
;  PRIOR FILING DATE:  1997-11-14
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-540-245A-2

  Query Match             16.2%;  Score 253.5;  DB 3;  Length 1525;
  Best Local Similarity   24.8%;  Pred. No. 1.6e-14;
  Matches   92;  Conservative   48;  Mismatches  100;  Indels  131;  Gaps   18;

Qy          1 CERQPCQHGATC-MPAGEYE-FQCLCRDGFKGDLCE-----HEENPC------------- 40
              |   ||:|| || :  || : | |:| |||:|: ||      |:| |
Db        957 CISNPCKHGGTCHLKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNY 1016

Qy         41 ---------------------QLREPCLHGGTC----QGTRCLCLPGFSGP--------- 66
                                   |   || |   |    :| :| | ||: |
Db       1017 TCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDC 1076

Qy         67 ---RCQQGSGHGIAESDWHL---EGSGG-----------------------NDA------ 91
                 :|: |:    | : :     ||  |                       | |
Db       1077 QDNKCKNGAHCTDAVNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRI 1136

Qy         92 --------PGQYGA---------YFHDDGFLAFPGHVFSRSLPEVPET-IELEVRTSTAS 133
                      ||  |          : : : :|  |      |    |:| | |:: |   |
Db       1137 NEPICQCLPGYQGEKCEKLVSVNFINKESYLQIP------SAKVRPQTNITLQIATDEDS 1190

Qy        134 GLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV-SEDPINDGEWHRVTAL 192
              |:||::|         || |:: |  | :   |  ||  |  : | : |||| :| |  |
Db       1191 GILLYKG--------DKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELL 1242

Qy        193 REGRRGSIQVDG--EELVSGRSPGPNVAVNAKGSVYIGGAP---DVATL--TGGRFSSGI 245
                 :  |: |||   ::::  |      :|    :|:|| |   :||:|    |:  :
Db       1243 ALDQSLSLSVDGGNPKIITNLS--KQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSF 1300

Qy        246 TGCVKNLVLHS 256
               ||::|| ::|
Db       1301 HGCIRNLYINS 1311

RESULT 14
US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER:  US/09/540,153
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  09/191,647
;  PRIOR FILING DATE:  1998-11-13
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-540-153-2

  Query Match             16.2%;  Score 253.5;  DB 3;  Length 1525;
  Best Local Similarity   24.8%;  Pred. No. 1.6e-14;
  Matches   92;  Conservative   48;  Mismatches  100;  Indels  131;  Gaps   18;

Qy          1 CERQPCQHGATC-MPAGEYE-FQCLCRDGFKGDLCE-----HEENPC------------- 40
              |   ||:|| || :  || : | |:| |||:|: ||      |:| |
Db        957 CISNPCKHGGTCHLKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNY 1016

Qy         41 ---------------------QLREPCLHGGTC----QGTRCLCLPGFSGP--------- 66
                                   |   || |   |    :| :| | ||: |
Db       1017 TCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDC 1076

Qy         67 ---RCQQGSGHGIAESDWHL---EGSGG-----------------------NDA------ 91
                 :|: |:    | : :     ||  |                       | |
Db       1077 QDNKCKNGAHCTDAVNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRI 1136

Qy         92 --------PGQYGA---------YFHDDGFLAFPGHVFSRSLPEVPET-IELEVRTSTAS 133
                      ||  |          : : : :|  |      |    |:| | |:: |   |
Db       1137 NEPICQCLPGYQGEKCEKLVSVNFINKESYLQIP------SAKVRPQTNITLQIATDEDS 1190

Qy        134 GLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV-SEDPINDGEWHRVTAL 192
              |:||::|         || |:: |  | :   |  ||  |  : | : |||| :| |  |
Db       1191 GILLYKG--------DKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELL 1242

Qy        193 REGRRGSIQVDG--EELVSGRSPGPNVAVNAKGSVYIGGAP---DVATL--TGGRFSSGI 245
                 :  |: |||   ::::  |      :|    :|:|| |   :||:|    |:  :
Db       1243 ALDQSLSLSVDGGNPKIITNLS--KQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSF 1300

Qy        246 TGCVKNLVLHS 256
               ||::|| ::|
Db       1301 HGCIRNLYINS 1311

RESULT 15
US-09-191-647-9
; Sequence 9, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
;  APPLICANT: Goodman, Corey
;  APPLICANT: Kid, Thomas
;  APPLICANT: Brose, Katja
;  APPLICANT: Tessier-Lavigne, Marc
;  TITLE OF INVENTION: Modulating Robo: Ligand Interactions
;  FILE REFERENCE: B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/191,647
;  CURRENT FILING DATE: 1998-11-13
;  EARLIER APPLICATION NUMBER: 60/065,544
;  EARLIER FILING DATE: 1997-11-14
;  EARLIER APPLICATION NUMBER: 60/081,057
;  EARLIER FILING DATE: 1998-04-07
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
```

```
;   LENGTH: 735
;   TYPE: PRT
;   ORGANISM: Caenorhabditis elegans
US-09-191-647-9

  Query Match             15.9%;  Score 249;  DB 3;  Length 735;
  Best Local Similarity   26.7%;  Pred. No. 1.6e-14;
  Matches   84;  Conservative   48;  Mismatches  115;  Indels   68;  Gaps   17;

Qy          5 PCQHGATCMPA-GEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTC----QGTRCLC 59
              ||::   |:|  | |   |:|  || |: ||   : |:  | | :||:|        |||
Db        320 PCENNGKCIPINGSY--SCMCSPGFTGNNCETNIDDCKNVE-CQNGGSCVDGILSYDCLC 376

Qy         60 LPGFSG------------------PRCQQ---GSGHGIA---ESDWHL---EGSGGNDAPGQ 94
               ||::|                  |||   | |  :|    ||:     ||  |     |
Db        377 RPGYAGQYCEIPPMMDMEYQKTDACQQSACGQGECVASQNSSDFTCKCHEGFSGPSCDRQ 436

Qy         95 YGAYFHDDG-FLAFPGHVFSRSLPEVPE-TIELEVRTSTASGLLLWQGVEVGEAGQGKDF 152
                  | : | :||          |   : || : :||::  |:||: | :         |
Db        437 MSVGFKNPGAYLALD--------PLASDGTITMTLRTTSKIGILLYYGDD--------HF 480

Qy        153 ISLGLQDGHLVFRYQLGSGEA-RLVSEDPINDGEWHRVTALREGRRGSIQVDGE--ELVS 209
              :|  | || :   | :|:  |  : |   :|||  ||::     |:  :|:|    ::|
Db        481 VSAELYDGRVKLVYYIGNFPASHMYSSVKVNDGLPHRISIRTSERKCFLQIDKNPVQIVE 540

Qy        210 GRSPGPNVAVNAKGSVYIGGAP-----DVATLTGGRFSSGITGCVKNLVLHSARPGAPPP 264
                     :    |  :|||| |     |       : |  : ||: :: ::
Db        541 NSGKSDQLITKGKEMLYIGGLPIEKSQDAKRRFHVKNSESLKGCISSITINEV------- 593

Qy        265 QPLDLQHRAQAGANT 279
               |::|| :|    ||
Db        594 -PINLQ-QALENVNT 606

Search completed: March  9, 2004, 17:23:33
Job time : 12.1326 secs
```

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:11:22 ; Search time 34.5822 Seconds
                                            (without alignments)
                                            2320.373 Million cell updates/sec

Title:           US-10-006-011A-9
Perfect score:   1566
Sequence:        1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        A_Geneseq_29Jan04:*
                  1:  geneseqp1980s:*
                  2:  geneseqp1990s:*
                  3:  geneseqp2000s:*
                  4:  geneseqp2001s:*
                  5:  geneseqp2002s:*
                  6:  geneseqp2003as:*
                  7:  geneseqp2003bs:*
                  8:  geneseqp2004s:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.

                                    SUMMARIES
                       %
Result               Query
  No.      Score     Match Length DB  ID                                  Description
-----------------------------------------------------------------------------------------
     1      1566    100.0   4391  6  AAE34390                             Aae34390 Human per
     2      1559     99.6   4393  4  AAB31889                             Aab31889 Amino aci
     3      1543     98.5   4436  4  ABG23265                             Abg23265 Novel hum
     4      1030     65.8    195  4  AAB31890                             Aab31890 Amino aci
     5     357.5     22.8   1931  6  ABU52400                             Abu52400 Human GPC
     6       343     21.9    238  4  AAU18145                             Aau18145 Novel hum
     7       343     21.9    238  4  AAU17011                             Aau17011 Human nov
     8       343     21.9    238  4  ABB10450                             Abb10450 Human cDN
     9       343     21.9    238  4  AAU19961                             Aau19961 Novel hum
    10       343     21.9    238  5  ABJ05772                             Abj05772 Novel hum
    11       343     21.9    238  5  ABP67037                             Abp67037 Human pol
    12       343     21.9    375  4  AAU07421                             Aau07421 Novel hum
    13       343     21.9    432  4  AAU18102                             Aau18102 Novel hum
    14       343     21.9    432  4  AAU16938                             Aau16938 Human nov
    15       343     21.9    432  4  ABB10233                             Abb10233 Human cDN
    16       343     21.9    432  4  AAU19905                             Aau19905 Novel hum
    17       343     21.9    432  5  ABJ05729                             Abj05729 Novel hum
    18       343     21.9    432  5  ABP66820                             Abp66820 Human pol
    19       343     21.9    463  7  ADB64904                             Adb64904 Human pro
    20       343     21.9    671  6  ABP58231                             Abp58231 Human cel
    21       343     21.9   1009  7  ADE28105                             Ade28105 Human NTR
    22     340.5     21.7    406  5  ABP41801                             Abp41801 Human ova
    23     338.5     21.6   2053  7  ADC39154                             Adc39154 Novel hum
    24     338.5     21.6   2143  7  ADC39164                             Adc39164 Novel hum
    25       334     21.3    416  2  AAY73993                             Aay73993 Human pro
```

```
    26       334     21.3   1565  7  ADC39156                             Adc39156 Novel hum
    27       334     21.3   1566  7  ADC39166                             Adc39166 Novel hum
    28       330     21.1    492  2  AAW26609                             Aaw26609 Human agr
    29       330     21.1    819  5  ABB72291                             Abb72291 Rat prote
    30       310     19.8   1544  7  ADC39162                             Adc39162 Novel hum
    31       303     19.3   1741  5  ABP43859                             Abp43859 Human mRN
    32       292     18.6   1298  4  ABB61948                             Abb61948 Drosophil
    33       292     18.6   1508  2  AAY27141                             Aay27141 Human sli
    34       292     18.6   1508  2  AAW96706                             Aaw96706 Protein s
    35       292     18.6   1508  2  AAY04138                             Aay04138 Human sli
    36       292     18.6   1534  2  AAW46966                             Aaw46966 Amino aci
    37       292     18.6   1534  2  AAY27144                             Aay27144 Human sli
    38       292     18.6   1534  2  AAW96707                             Aaw96707 Protein s
    39       292     18.6   1534  2  AAY04139                             Aay04139 Human sli
    40       292     18.6   1534  7  ADE54473                             Ade54473 Human Pro
    41     290.5     18.6    210  4  AAU19782                             Aau19782 Human nov
    42     290.5     18.6    210  5  ABP48002                             Abp48002 Human pol
    43     290.5     18.6    210  7  ADC10964                             Adc10964 Human ext
    44       281     17.9   4072  4  ABB63614                             Abb63614 Drosophil
    45     279.5     17.8    152  5  ABP56528                             Abp56528 BM hepara

                                   ALIGNMENTS

RESULT 1
AAE34390
ID   AAE34390 standard; protein; 4391 AA.
XX
AC   AAE34390;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Human perlecan protein.
XX
KW   Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX
OS   Homo sapiens.
XX
PN   WO200295415-A2.
XX
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-EP005612.
XX
PR   23-MAY-2001; 2001GB-00012626.
XX
PA   (OSTE-) OSTEOMETER BIO TECH AS.
XX
PI   Christgau S,  Henriksen DB,  Cloos PAC;
XX
DR   WPI; 2003-140389/13.
XX
PT   An assay for the diagnosis or assessment of the severity of
PT   osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
PT   or optically inverted protein in a sample.
XX
PS   Disclosure; Page 46-67; 106pp; English.
XX
CC   The invention relates to an assay for the diagnosis or assessment of the
CC   severity of osteoarthritis or rheumatoid arthritis. The assay involves
CC   measuring (in a biological sample) the amount or presence of an
CC   isomerised or optically inverted protein or one or more isomerised or
CC   optically inverted fragments from proteins such as perlecan, biglycan,
CC   decorin, fibrillin-1 or protocadherin. The assay is useful for the
CC   diagnosis or assessment of the severity of osteoarthritis or rheumatoid
CC   arthritis. The present sequence is human perlecan protein
XX
SQ   Sequence 4391 AA;

  Query Match             100.0%;  Score 1566;  DB 6;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-110;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4108 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4167

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4168 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4227

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4228 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 4287

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4288 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 4347

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||||||||||||||||||
Db       4348 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391

RESULT 2
AAB31889
ID   AAB31889 standard; protein; 4393 AA.
XX
AC   AAB31889;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a human protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 138-152; 209pp; French.
XX
CC   The present sequence represents a human protein, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 4393 AA;

  Query Match             99.6%;  Score 1559;  DB 4;  Length 4393;
  Best Local Similarity   99.3%;  Pred. No. 2.8e-109;
  Matches  282;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       4110 CERQPCQHGATCMPAGEYEFQCLCRDGIKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4169

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4170 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4229

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4230 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 4289

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db       4290 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSIYIGGAPDVATLTGGR 4349

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||||||||||||||||||
Db       4350 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4393

RESULT 3
ABG23265
ID   ABG23265 standard; protein; 4436 AA.
XX
AC   ABG23265;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23256.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87452.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 53624; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
```

```
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4436 AA;

  Query Match             98.5%;  Score 1543;  DB 4;  Length 4436;
  Best Local Similarity   96.2%;  Pred. No. 4.7e-108;
  Matches  282;  Conservative    0;  Mismatches    1;  Indels   10;  Gaps    1;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4143 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4202

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4203 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4262

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVF----------RYQLGS 170
              ||||||||||||||||||||||||||||||||||||||||||||          ||||||
Db       4263 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFSSYFLGLDDCRYQLGS 4322

Qy        171 GEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGA 230
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       4323 GEARLVSEDPFNDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGA 4382

Qy        231 PDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 283
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4383 PDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 4435

RESULT 4
AAB31890
ID   AAB31890 standard; protein; 195 AA.
XX
AC   AAB31890;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of the C-terminal of the human perlecan protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;    99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
```

```
XX
DR   WPI; 2001-159475/16.
DR   N-PSDB; AAF54728.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 152-153; 209pp; French.
XX
CC   The present sequence represents a human polypeptide, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 195 AA;

  Query Match             65.8%;  Score 1030;  DB 4;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-70;
  Matches  195;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         90 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 60

Qy        150 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 120

Qy        210 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 180

Qy        270 QHRAQAGANTRPCPS 284
              |||||||||||||||
Db        181 QHRAQAGANTRPCPS 195

RESULT 5
ABU52400
ID   ABU52400 standard; protein; 1931 AA.
XX
AC   ABU52400;
XX
DT   03-MAR-2003  (first entry)
XX
DE   Human GPCR related protein NOV40a.
XX
KW   Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200279398-A2.
XX
PD   10-OCT-2002.
XX
PF   08-MAR-2002; 2002WO-US007355.
XX
PR   08-MAR-2001; 2001US-0274194P.
PR   08-MAR-2001; 2001US-0274281P.
PR   08-MAR-2001; 2001US-0274322P.
PR   09-MAR-2001; 2001US-0274849P.
```

```
PR   13-MAR-2001; 2001US-0275578P.
PR   13-MAR-2001; 2001US-0275579P.
PR   13-MAR-2001; 2001US-0275601P.
PR   14-MAR-2001; 2001US-0276000P.
PR   16-MAR-2001; 2001US-0276776P.
PR   19-MAR-2001; 2001US-0276994P.
PR   20-MAR-2001; 2001US-0277239P.
PR   20-MAR-2001; 2001US-0277327P.
PR   20-MAR-2001; 2001US-0277338P.
PR   21-MAR-2001; 2001US-0277791P.
PR   22-MAR-2001; 2001US-0277833P.
PR   23-MAR-2001; 2001US-0278152P.
PR   26-MAR-2001; 2001US-0278894P.
PR   27-MAR-2001; 2001US-0278999P.
PR   27-MAR-2001; 2001US-0279036P.
PR   30-MAR-2001; 2001US-0280233P.
PR   02-APR-2001; 2001US-0280802P.
PR   02-MAY-2001; 2001US-0288052P.
PR   02-MAY-2001; 2001US-0288066P.
PR   02-MAY-2001; 2001US-0288228P.
PR   17-MAY-2001; 2001US-0291766P.
PR   07-JUN-2001; 2001US-0296693P.
PR   08-JUN-2001; 2001US-0296856P.
PR   05-JUL-2001; 2001US-0303230P.
PR   05-JUL-2001; 2001US-0303237P.
PR   08-AUG-2001; 2001US-0310913P.
PR   13-AUG-2001; 2001US-0311978P.
PR   14-AUG-2001; 2001US-0312191P.
PR   16-AUG-2001; 2001US-0312916P.
PR   17-AUG-2001; 2001US-0313182P.
PR   20-AUG-2001; 2001US-0313626P.
PR   21-AUG-2001; 2001US-0314018P.
PR   27-AUG-2001; 2001US-0315227P.
PR   10-SEP-2001; 2001US-0318403P.
PR   10-SEP-2001; 2001US-0318510P.
PR   14-SEP-2001; 2001US-0322296P.
PR   14-SEP-2001; 2001US-0322360P.
PR   27-SEP-2001; 2001US-0325378P.
PR   09-NOV-2001; 2001US-0332486P.
PR   09-NOV-2001; 2001US-0345399P.
PR   07-MAR-2002; 2002US-00094886.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Kekuda R,  Tchernev VT,  Liu X,  Spytek KA,  Patturajan M;
PI   Burgess CE,  Vernet CAM,  Li L,  Gorman L,  Malyankar UM,  Boldog FL;
PI   Guo X,  Shenoy SG,  Padigaru M,  Taupier RJ,  Miller CE,  Casman SJ;
PI   Pena CEA,  Gangolli EA,  Gusev V,  Smithson G,  Zerhusen BD,  Gerlach V;
PI   Pochart PF,  Fernandes ER,  Shimkets RA,  Rastelli L,  Spaderna SK;
PI   Larochelle WJ,  Zhong M,  Khramtsov NV,  Voss EZ,  Herrmann JL;
XX
DR   WPI; 2003-058423/05.
DR   N-PSDB; ABX70491.
XX
PT   NOVX polypeptides and polynucleotides, useful for treating a syndrome
PT   related to a human disease associated with the NOVX polypeptide e.g.,
PT   cancer.
XX
PS   Claim 1; Page 223; 413pp; English.
XX
CC   The present invention relates to the isolation of novel human
CC   polypeptides referred to as NOVX (NOV1-NOV44), variants of these
CC   proteins, and the polynucleotide sequences encoding them. The NOVX
CC   proteins of the invention are G-protein coupled receptor (GPCR) related
CC   proteins. The sequences of the invention are useful in the manufacture of
CC   a medicament for treating a syndrome related to a human disease
CC   associated with the polypeptides e.g. cancer. ABU52311-ABU52408 represent
CC   the human NOVX proteins of the invention
XX
SQ   Sequence 1931 AA;

  Query Match             22.8%;  Score 357.5;  DB 6;  Length 1931;
  Best Local Similarity   35.4%;  Pred. No. 3.1e-18;
```

```
  Matches   91;  Conservative   29;  Mismatches  110;  Indels   27;  Gaps    8;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
              |   ||  || |       | | |  |  |  |  |::|||   ||     |:     |
Db       1439 CLPNPCHGGAPCQNLEAGRFHCQCPPGRVGPTCADEKSPCQ-PNPCHGAAPCRVLPEGGA 1497

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
              :| |  |  |  ||  |               | |  | : | |:    |   | | |:|
Db       1498 QCECPLGREGTFCQTAS---------------GQDGSGPFLADFNGFSHLELRGLHTFAR 1542

Qy        115 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 174
               | |    :|:       |||||: | :    |:| ||:|| |:|  | ||| || | |
Db       1543 DLGE-KMALEVVFLARGPSGLLLYNGQKTD--GKG-DFVSLALRDRRLEFRYDLGKGAAV 1598

Qy        175 LVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVA 234
              : | :|:  | | ||:  | ||:|:::|     | | || |:  :| |  :|:||||| :
Db       1599 IRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFS 1658

Qy        235 TLT-GGRFSSGITGCVK 250
               |      |||  | ::
Db       1659 KLARAAAVSSGFDGAIQ 1675


RESULT 6
AAU18145
ID   AAU18145 standard; protein; 238 AA.
XX
AC   AAU18145;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Novel human uterine motility-association polypeptide #52.
XX
KW   Human; uterine motility-association disorder; uterus; pregnancy; labour;
KW   menstrual cycle; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200155201-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001317.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
```

```
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488777/53.
DR   N-PSDB; AAS28987.
XX
PT   New nucleic acid molecules encoding 49 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 121; 524pp; English.
XX
CC   The present invention relates to the isolation of novel human uterine
CC   motility-association polypeptides, and cDNA (AAS28936-AAS28994) and
CC   genomic sequences encoding for these polypeptides. The sequences of the
CC   invention are useful in the diagnosis, treatment, prevention and/or
CC   prognosis of diseases associated with uterine motility such as pregnancy
CC   and labour, and menstrual disorders. The polynucleotide sequences of the
CC   invention are also useful in gene therapy. AAU18094-AAU18152 represent
CC   novel human uterine motility-association polypeptides. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 238 AA;

  Query Match             21.9%;  Score 343;  DB 4;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.6e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db          11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db          57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105
```

```
Qy          146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                     |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db          106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy          206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                   :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  |  :
Db          166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy          266 PLDLQHRAQAGANTRPC 282
                 : |   |  | |   |
Db          219 HISLVEDAVDGKNINTC 235


RESULT 7
AAU17011
ID   AAU17011 standard; protein; 238 AA.
XX
AC   AAU17011;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, SEQ ID 252.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155441-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001320.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
```

```
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476222/51.
DR   N-PSDB; AAS26916.
XX
PT   Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT   diagnose diseases or disorders associated with aberrant expression or
PT   activity of polypeptides, for treating blood clotting disorder,
PT   hemophilia.
XX
PS   Claim 11; SEQ ID NO 252; 601pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The

  Query Match             21.9%;  Score 343;  DB 4;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.6e-18;
```

```
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy           34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
                |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db           11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------ 56

Qy           90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
                : |      |  | : |      |   |||       : :  :|:   |||||:|
Db           57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy          146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                     |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db          106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy          206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                   :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db          166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy          266 PLDLQHRAQAGANTRPC 282
                 : |   |  | |   |
Db          219 HISLVEDAVDGKNINTC 235


RESULT 8
ABB10450
ID   ABB10450 standard; protein; 238 AA.
XX
AC   ABB10450;
XX
DT   10-JAN-2002  (first entry)
XX
DE   Human cDNA SEQ ID NO: 758.
XX
KW   Human; gene therapy; neural disorder; immune system disorder;
KW   muscular disorder; reproductive disorder; gastrointestinal disorder;
KW   pulmonary disorder; cardiovascular disorder; renal disorder;
KW   proliferative disorder; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200154474-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001349.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
```

```
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476161/51.
DR   N-PSDB; ABA06672.
XX
PT   Isolated nucleic acid molecule encoding an inflammation-associated
PT   polypeptide is used in preventing, treating or ameliorating a medical
PT   condition.
XX
PS   Claim 11; SEQ ID NO 758; 859pp + Sequence Listing; English.
XX
CC   The present invention provides human cDNAs, proteins and related genomic
CC   DNAs. These can be used in the treatment of neural, immune system,
CC   muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC   renal and proliferative disorders and inflammation. The present sequence
CC   is a protein of the invention
XX
SQ   Sequence 238 AA;

  Query Match             21.9%;  Score 343;  DB 4;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.6e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy           34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
                |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db           11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy           90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
                : |      |  | : |      |   |||       : :  :|:   |||||:|
Db           57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy          146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                     |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db          106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165
```

```
Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        219 HISLVEDAVDGKNINTC 235


RESULT 9
AAU19961
ID   AAU19961 standard; protein; 238 AA.
XX
AC   AAU19961;
XX
DT   06-DEC-2001  (first entry)
XX
DE   Novel human calcium-binding protein #70.
XX
KW   Human; calcium-binding protein; calcium flux; neurological disease;
KW   immune dysfunction; digestive disorder; neoplastic disease;
KW   blood disorder; infectious disease; gene therapy; immunosuppressive;
KW   antiarthritic; cytostatic; vasotropic; antibacterial; nootropic;
KW   virucide.
XX
OS   Homo sapiens.
XX
PN   WO200155304-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001302.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
```

```
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465568/50.
DR   N-PSDB; AAS31646.
XX
PT   Isolated nucleic acid molecule encoding a calcium-binding protein is used
PT   in preventing, treating or ameliorating a medical condition.
XX
PS   Claim 11; SEQ ID NO 158; 542pp; English.
XX
CC   The present invention relates to the isolation of novel human calcium-
CC   binding proteins, and cDNA (AAS31577-AAS31654) and genomic sequences
CC   encoding for these proteins. The sequences of the invention are useful in
CC   the diagnosis, prevention and/or prognosis of diseases associated with
CC   aberrant calcium flux. Such disorders include neurological diseases (e.g.
CC   amylotrophic lateral sclerosis, ALS), immune dysfunction (e.g. severe
CC   combined immunodeficiency, SCID), digestive disorders (e.g. irritable
CC   bowel syndrome, IBS), neoplastic disease (e.g. cancer), blood disorders
CC   (e.g. haemophilia), and/or infectious disease (e.g. acquired
CC   immunodeficiency syndrome, AIDS). The novel calcium-binding proteins are
CC   also useful as screening tools to identify antagonists and/or agonists
CC   that may enhance or inhibit activities mediated by calcium-binding
CC   proteins. The polynucleotides of the invention are also useful in gene
CC   therapy. AAU19892-AAU19969 represent the novel human calcium-binding
CC   proteins. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 238 AA;

  Query Match             21.9%;  Score 343;  DB 4;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.6e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db          11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db          57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         219 HISLVEDAVDGKNINTC 235

RESULT 10
ABJ05772
ID   ABJ05772 standard; protein; 238 AA.
XX
AC   ABJ05772;
XX
DT   14-NOV-2002  (first entry)
XX
DE   Novel human protein SEQ ID No 121.
XX
KW   Immunostimulant; antirheumatic; antiarthritic; neuroprotective;
KW   antiallergic; antidiabetic; antiasthmatic; antiinflammatory; nootropic;
KW   immunosuppressive; anticoagulant; thrombolytic; antiatherosclerotic;
KW   cytostatic; nephrotropic; antiparkinsonian; gynecological; virucide;
KW   antibacterial; antiarrhythmic; fungicide; HCFAT05; HWAAE95; HTNBM01;
KW   immunodeficiency; autoimmune disorder; allergic reaction; cardiovascular;
KW   inflammatory condition; graft-versus-host disease; reproductive system;
KW   blood-related disorder; hyperproliferative; endocrine; neurological;
KW   respiratory; renal; infectious disease; gastrointestinal; gene therapy;
KW   neuronal growth; neuronal disorder; neuro-degenerative condition;
KW   keratinocyte growth; human.
XX
OS   Homo sapiens.
XX
PN   US2002086330-A1.
XX
PD   04-JUL-2002.
XX
PF   17-JAN-2001; 2001US-00764893.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
```

```
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2002-665432/71.
XX
PT   Novel polypeptide useful for diagnosis, prognosis, prevention, and
PT   treatment of immune, hyperproliferative, renal, respiratory,
PT   cardiovascular, reproductive, endocrine, gastrointestinal and
PT   neurological disorders.
XX
PS   Disclosure; Page 287-288; 335pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a sequence at
CC   least 90% identical to a full length protein sequence selected from 55
CC   sequences given in the specification such as a sequence of 163, 74 or 140
CC   amino acids fully defined in the specification, or the encoding sequence
CC   contained in 49 cDNA clones given in specification e.g. HCFAT05, HWAAE95
CC   or HTNBM01. The protein and its encoding nucleic acid are useful for
CC   diagnosing a pathological condition or susceptibility to a pathological
CC   condition in a subject and for preventing, treating or ameliorating a
CC   medical condition. The protein, its encoding nucleic acid and an isolated
CC   antibody that can bind to the protein are useful in treating, preventing,
CC   diagnosing and/or prognosing immunodeficiencies, autoimmune disorders,
CC   allergic reactions and conditions, inflammatory conditions, graft-versus-
CC   host disease, blood-related disorders, hyperproliferative disorders,
CC   renal disorders, cardiovascular disorders, respiratory disorders,
CC   neurological disorders, endocrine disorders, reproductive system
CC   disorders, infectious diseases, and gastrointestinal disorders. The
CC   protein of the invention is useful to stimulate neuronal growth and to
CC   treat, prevent, and/or diagnose neuronal damage which occurs in certain
CC   neuronal disorders or neuro-degenerative conditions, for stimulating
CC   keratinocyte growth, to prevent hair loss, to modulate mammalian
CC   characteristics such as body height, weight, hair color, and to increase
CC   or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrate, vitamins, minerals, cofactors or other nutritional
CC   components. The nucleic acid of the invention can be used in gene
CC   therapy. This sequence represents a novel human protein of the invention
XX
SQ   Sequence 238 AA;

  Query Match             21.9%;  Score 343;  DB 5;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.6e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db          11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db          57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         219 HISLVEDAVDGKNINTC 235

RESULT 11
ABP67037
ID   ABP67037 standard; protein; 238 AA.
XX
AC   ABP67037;
XX
DT   09-DEC-2002  (first entry)
XX
DE   Human polypeptide SEQ ID NO 758.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   US2002090672-A1.
XX
PD   11-JUL-2002.
XX
PF   17-JAN-2001; 2001US-00764853.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
```

```
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2002-681727/73.
DR   N-PSDB; ABV84009.
XX
PT   Novel polypeptide useful for diagnosis, prognosis, prevention, and
PT   treatment of immune, hyperproliferative, renal, respiratory,
PT   cardiovascular, reproductive, endocrine, gastrointestinal and
PT   neurological disorders.
XX
PS   Claim 11; SEQ ID NO 758; 369pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABV83682-ABV84101) and proteins
CC   (ABP66710-ABP67129) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 238 AA;

  Query Match             21.9%;  Score 343;  DB 5;  Length 238;
  Best Local Similarity   33.5%;  Pred. No. 3.6e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db          11 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 56

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db          57 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 105

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         106 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         166 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 218

Qy         266 PLDLQHRAQAGANTRPC 282
```

```
                : |   |  | |   |
Db         219 HISLVEDAVDGKNINTC 235

RESULT 12
AAU07421
ID   AAU07421 standard; protein; 375 AA.
XX
AC   AAU07421;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human extracellular matrix (ECM) protein #4.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW   extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW   nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
KW   wound healing; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200179253-A1.
XX
PD   25-OCT-2001.
XX
PF   11-APR-2001; 2001WO-US011643.
XX
PR   18-APR-2000; 2000US-0198123P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Fiscella M,  Shi Y,  Ebner R,  Ruben SM;
XX
DR   WPI; 2001-611720/70.
DR   N-PSDB; AAS13846.
XX
PT   New nucleic acids encoding extracellular matrix polypeptides, for
PT   diagnosing, treating, preventing or ameliorating human disorders and
PT   disease, such as, autoimmune, hyperproliferative or cardiovascular
PT   disorders.
XX
PS   Claim 1; Page 297-298; 308pp; English.
XX
CC   The invention relates to novel isolated polynucleotides (I) encoding
CC   extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
CC   (I) are used to prevent, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC   are also used in diagnosing a pathological condition or susceptibility to
CC   a pathological condition. The antibodies to the polypeptides can also be
CC   used in alleviating symptoms associated with the disorders and in
CC   diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC   immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC   include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. The polypeptides can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. The present sequence
CC   represents the amino acid sequence of novel human extracellular matrix
CC   (ECM) protein #4
XX
SQ   Sequence 375 AA;

  Query Match             21.9%;  Score 343;  DB 4;  Length 375;
```

```
  Best Local Similarity   33.5%;  Pred. No. 6.1e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         148 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 193

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         194 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 242

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         243 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 302

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         303 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 355

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         356 HISLVEDAVDGKNINTC 372


RESULT 13
AAU18102
ID   AAU18102 standard; protein; 432 AA.
XX
AC   AAU18102;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Novel human uterine motility-association polypeptide #9.
XX
KW   Human; uterine motility-association disorder; uterus; pregnancy; labour;
KW   menstrual cycle; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200155201-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001317.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
```

```
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488777/53.
DR   N-PSDB; AAS28944.
XX
PT   New nucleic acid molecules encoding 49 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 78; 524pp; English.
XX
CC   The present invention relates to the isolation of novel human uterine
CC   motility-association polypeptides, and cDNA (AAS28936-AAS28994) and
CC   genomic sequences encoding for these polypeptides. The sequences of the
CC   invention are useful in the diagnosis, treatment, prevention and/or
CC   prognosis of diseases associated with uterine motility such as pregnancy
CC   and labour, and menstrual disorders. The polynucleotide sequences of the
CC   invention are also useful in gene therapy. AAU18094-AAU18152 represent
CC   novel human uterine motility-association polypeptides. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 432 AA;

  Query Match             21.9%;  Score 343;  DB 4;  Length 432;
  Best Local Similarity   33.5%;  Pred. No. 7.1e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db         300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  |  :
Db         360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         413 HISLVEDAVDGKNINTC 429

RESULT 14
AAU16938
ID   AAU16938 standard; protein; 432 AA.
XX
AC   AAU16938;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, SEQ ID 179.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155441-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001320.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
```

```
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476222/51.
DR   N-PSDB; AAS26843.
XX
PT   Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT   diagnose diseases or disorders associated with aberrant expression or
PT   activity of polypeptides, for treating blood clotting disorder,
PT   hemophilia.
XX
PS   Claim 11; SEQ ID NO 179; 601pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The

  Query Match            21.9%;  Score 343;  DB 4;  Length 432;
```

```
  Best Local Similarity   33.5%;  Pred. No. 7.1e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy           34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
                |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db          205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250

Qy           90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
                : |      |  | : |      |   |||       : :  :|:   |||||:|
Db          251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy          146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                     |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db          300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy          206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                   :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db          360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy          266 PLDLQHRAQAGANTRPC 282
                 : |   |  | |   |
Db          413 HISLVEDAVDGKNINTC 429


RESULT 15
ABB10233
ID   ABB10233 standard; protein; 432 AA.
XX
AC   ABB10233;
XX
DT   10-JAN-2002  (first entry)
XX
DE   Human cDNA SEQ ID NO: 541.
XX
KW   Human; gene therapy; neural disorder; immune system disorder;
KW   muscular disorder; reproductive disorder; gastrointestinal disorder;
KW   pulmonary disorder; cardiovascular disorder; renal disorder;
KW   proliferative disorder; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200154474-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001349.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
```

```
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476161/51.
DR   N-PSDB; ABA06455.
XX
PT   Isolated nucleic acid molecule encoding an inflammation-associated
PT   polypeptide is used in preventing, treating or ameliorating a medical
PT   condition.
XX
PS   Claim 11; SEQ ID NO 541; 859pp + Sequence Listing; English.
XX
CC   The present invention provides human cDNAs, proteins and related genomic
CC   DNAs. These can be used in the treatment of neural, immune system,
CC   muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC   renal and proliferative disorders and inflammation. The present sequence
CC   is a protein of the invention
XX
SQ   Sequence 432 AA;

  Query Match             21.9%;  Score 343;  DB 4;  Length 432;
  Best Local Similarity   33.5%;  Pred. No. 7.1e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy          34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
               |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db         205 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 250

Qy          90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
               : |      |  | : |      |   |||       : :  :|:   |||||:|
Db         251 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 299

Qy         146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                    |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
```

```
Db         300 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 359

Qy         206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                  :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  |  :
Db         360 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 412

Qy         266 PLDLQHRAQAGANTRPC 282
                : |   |  | |   |
Db         413 HISLVEDAVDGKNINTC 429

Search completed: March  9, 2004, 17:19:03
Job time : 35.5822 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:11:52 ; Search time 17.0517 Seconds
                                           (without alignments)
                                           2152.832 Million cell updates/sec

Title:          US-10-006-011A-3
Perfect score:  3825
Sequence:       1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                  SUMMARIES
                     %
Result            Query
  No.    Score    Match Length DB  ID                   Description
------------------------------------------------------------------------------
   1     3825    100.0   4391  1  PGBM_HUMAN           P98160 homo sapien
   2   3389.5     88.6   3707  1  PGBM_MOUSE           Q05793 mus musculu
   3      859     22.5   1959  1  AGRI_RAT             P25304 rattus norv
   4    796.5     20.8   1955  1  AGRI_CHICK           P31696 gallus gall
   5      667     17.4   1328  1  AGRI_DISOM           Q90404 discopyge o
   6      511     13.4   3375  1  UN52_CAEEL           Q06561 caenorhabdi
   7      441     11.5   3313  1  CLR3_RAT             O88278 rattus norv
   8      421     11.0   1715  1  NX2A_RAT             Q63374 rattus norv
   9    420.5     11.0   1712  1  NX2A_HUMAN           Q9p2s2 homo sapien
  10    412.5     10.8   1477  1  NX1A_HUMAN           Q9ulb1 homo sapien
  11    405.5     10.6   1530  1  NX1A_BOVIN           Q28146 bos taurus
  12    403.5     10.5   1514  1  NX1A_RAT             Q63372 rattus norv
  13      403     10.5   3301  1  CLR3_MOUSE           Q91zi0 mus musculu
  14    397.5     10.4   1529  1  SLT2_HUMAN           O94813 homo sapien
  15    397.5     10.4   3312  1  CLR3_HUMAN           Q9nyq7 homo sapien
  16      396     10.4   1541  1  NX3A_HUMAN           Q9y4c0 homo sapien
  17      394     10.3   1363  1  NX1A_CHICK           Q9ddd0 gallus gall
  18      389     10.2   2144  1  CLR2_RAT             Q9qyp2 rattus norv
  19      388     10.1   5147  1  FAT_DROME            P33450 drosophila
  20      386     10.1   3579  1  STAN_DROME           Q9v5n8 drosophila
  21      382     10.0   2923  1  CLR2_HUMAN           Q9hcu4 homo sapien
  22      367      9.6   1578  1  NX3A_RAT             Q07310 rattus norv
  23      364      9.5   2920  1  CLR2_MOUSE           Q9r0m0 mus musculu
  24      357      9.3   3110  1  LMA2_HUMAN           P24043 homo sapien
  25    354.5      9.3   1310  1  CTA4_MOUSE           Q99p47 mus musculu
  26    353.5      9.2   3084  1  LMA1_MOUSE           P19137 mus musculu
  27      353      9.2   3075  1  LMA1_HUMAN           P25391 homo sapien
  28      351      9.2   3097  1  CADN_DROME           O15943 drosophila
  29      347      9.1   3106  1  LMA2_MOUSE           Q60675 mus musculu
  30      333      8.7   1504  1  SLIT_DROME           P24014 drosophila
  31    328.5      8.6   2471  1  NTC2_HUMAN           Q04721 homo sapien
  32      328      8.6   4705  1  FAT2_DROME           Q9vw71 drosophila
  33      326      8.5   3672  1  LML2_CAEEL           Q21313 caenorhabdi
  34      325      8.5   3014  1  CLR1_HUMAN           Q9nyq6 homo sapien
  35    322.5      8.4   4590  1  FATH_HUMAN           Q14517 homo sapien
  36      319      8.3   2471  1  NTC2_RAT             Q9qw30 rattus norv
  37      316      8.3   2321  1  NTC3_HUMAN           Q9um47 homo sapien
  38    314.5      8.2   2470  1  NTC2_MOUSE           O35516 mus musculu
  39      312      8.2   2003  1  NTC4_HUMAN           Q99466 homo sapien
  40    311.5      8.1   2319  1  NTC3_RAT             Q9r172 rattus norv
  41      311      8.1   2318  1  NTC3_MOUSE           Q61982 mus musculu
  42      311      8.1   3034  1  CLR1_MOUSE           O35161 mus musculu
  43    310.5      8.1   1308  1  CTA4_HUMAN           Q9c0a0 homo sapien
  44    303.5      7.9   1331  1  CTA2_HUMAN           Q9uhc6 homo sapien
  45    302.5      7.9   1219  1  JAG1_RAT             Q63722 rattus norv

                                  ALIGNMENTS

RESULT 1
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4391 AA.
AC   P98160; Q16287; Q9H3V5;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Skin;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX   MEDLINE=20553141; PubMed=11101850;
RA   Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA   Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA   Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA   Hentati F., Fontaine B.;
RT   "Perlecan, the major proteoglycan of basement membranes, is altered in
RT   patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL   Nat. Genet. 26:480-483(2000).
RN   [4]
RP   SEQUENCE OF 1016-1470 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [5]
RP   SEQUENCE OF 890-1396 FROM N.A.
RC   TISSUE=Fibrosarcoma;
```

```
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [6]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN   [7]
RP   CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX   MEDLINE=22660472; PubMed=12754519;
RA   Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT   "Identification and quantification of N-linked glycoproteins using
RT   hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL   Nat. Biotechnol. 21:660-666(2003).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC       syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC       characterized by permanent myotonia (prolonged failure of muscle
CC       relaxation) and skeletal dysplasia, resulting in reduced stature,
CC       kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; AL445795; CAC18534.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   PIR; A38096; A38096.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
DR   MIM; 142461; -.
DR   MIM; 255800; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00180; EGF_Lam; 12.
DR   SMART; SM00409; IG; 22.
DR   SMART; SM00408; IGc2; 21.
DR   SMART; SM00406; IGv; 7.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 22.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4391       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      198    235       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      284    320       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      324    360       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      367    404       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      405    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1955       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1956   2051       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2052   2151       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2152   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2340       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2341   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2533       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2534   2629       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2630   2726       IG-LIKE C2-TYPE 12.
```

```
  Query Match             100.0%;  Score 3825;  DB 1;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-233;
  Matches  705;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3687 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3746

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3747 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3806

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3807 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3866

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3867 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3926

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3927 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3987 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4046

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4047 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4106

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4107 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4166

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4167 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4226

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4227 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4286

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4287 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 4346

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4347 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
PGBM_MOUSE
ID   PGBM_MOUSE     STANDARD;      PRT;  3707 AA.
AC   Q05793;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=92078153; PubMed=1744087;
RA   Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "The complete sequence of perlecan, a basement membrane heparan
RT   sulfate proteoglycan, reveals extensive similarity with laminin A
RT   chain, low density lipoprotein-receptor, and the neural cell adhesion
RT   molecule.";
RL   J. Biol. Chem. 266:22939-22947(1991).
RN   [2]
RP   SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89034110; PubMed=2972708;
RA   Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "Identification of cDNA clones encoding different domains of the
RT   basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 263:16379-16387(1988).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M77174; AAA39911.1; -.
DR   EMBL; J04054; AAA39899.1; -.
DR   EMBL; J04055; AAA39912.1; -.
DR   PIR; S18252; S18252.
DR   PDB; 1GL4; 28-NOV-01.
DR   MGD; MGI:96257; Hspg2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0008104; P:protein localization; IMP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 15.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 7.
DR   SMART; SM00408; IGc2; 14.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
```

```
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 15.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; 3D-structure.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   3707       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      195    234       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      281    319       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      320    359       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      360    403       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      404    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1954       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1955   2049       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2050   2148       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2149   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2343       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2344   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2532       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2533   2619       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2620   2720       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2721   2809       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2810   2895       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2896   2980       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2984   3162       LAMININ G-LIKE 1.
FT   DOMAIN     3163   3241       EGF-LIKE.
FT   DOMAIN     3245   3425       LAMININ G-LIKE 2.
FT   DOMAIN     3518   3705       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       3615   3617       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    332    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    368    381       BY SIMILARITY.
```

```
FT   DISULFID    375    394       BY SIMILARITY.
FT   DISULFID    388    403       BY SIMILARITY.
FT   DISULFID    428    479       BY SIMILARITY.
FT   DISULFID    764    773       BY SIMILARITY.
FT   DISULFID    766    780       BY SIMILARITY.
FT   DISULFID    783    792       BY SIMILARITY.
FT   DISULFID    795    811       BY SIMILARITY.
FT   DISULFID    814    829       BY SIMILARITY.
FT   DISULFID    816    839       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    869       BY SIMILARITY.
FT   DISULFID   1159   1168       BY SIMILARITY.
FT   DISULFID   1161   1175       BY SIMILARITY.
FT   DISULFID   1178   1187       BY SIMILARITY.
FT   DISULFID   1190   1206       BY SIMILARITY.
FT   DISULFID   1209   1224       BY SIMILARITY.
FT   DISULFID   1211   1234       BY SIMILARITY.
FT   DISULFID   1237   1246       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1275   1287       BY SIMILARITY.
FT   DISULFID   1277   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1307   1322       BY SIMILARITY.
FT   DISULFID   1563   1572       BY SIMILARITY.
FT   DISULFID   1565   1579       BY SIMILARITY.
FT   DISULFID   1582   1591       BY SIMILARITY.
FT   DISULFID   1594   1610       BY SIMILARITY.
FT   DISULFID   1613   1628       BY SIMILARITY.
FT   DISULFID   1615   1638       BY SIMILARITY.
FT   DISULFID   1641   1650       BY SIMILARITY.
FT   DISULFID   1653   1668       BY SIMILARITY.
FT   DISULFID   1792   1839       BY SIMILARITY.
FT   DISULFID   1886   1932       BY SIMILARITY.
FT   DISULFID   1976   2021       BY SIMILARITY.
FT   DISULFID   2073   2118       BY SIMILARITY.
FT   DISULFID   2170   2215       BY SIMILARITY.
FT   DISULFID   2268   2313       BY SIMILARITY.

  Query Match             88.6%;  Score 3389.5;  DB 1;  Length 3707;
  Best Local Similarity   88.9%;  Pred. No. 3.6e-206;
  Matches  627;  Conservative   33;  Mismatches   40;  Indels    5;  Gaps    3;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db       3008 EIKITFRPDSADGMLLYNGQKR---SPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3064

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||| ||||||||||||||||||:|||||||||||||||||||||||||||||||
Db       3065 RHPTPLALGQFHTVTLLRSLTQGSLIVGNLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3124

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||:|||||||||||||||||:||| |||||||||:|||||||||| ||||||
Db       3125 IPKAGLSSGFVGCVRELRIQGEEIVFHDVNLTTHGISHCPTCQDRPCQNGGQCQDSESSS 3184

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              | |||||||| :     :     | :   |||||||||||||||||||||||:|||||||
Db       3185 YTCVCPAGFTAAAVNIRKPCTATP-SLWADATCVNRPDGRGYTCRCHLGRSGVRCEEGVT 3243

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              |||||:|||||||||||||||||||||||||||| |:|:|||||||||||||||||||||
Db       3244 VTTPSMSGAGSYLALPALTNTHHELRLDVEFKPLEPNGILLFSGGKSGPVEDFVSLAMVG 3303

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              |||||||||||||||||| ||||||||||||||||||||||||:||||||||||||||||
Db       3304 GHLEFRYELGSGLAVLRSHEPLALGRWHRVSAERLNKDGSLRVDGGRPVLRSSPGKSQGL 3363

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||| ||||||||||| ||||||||||||||||||||||||:|||||||
Db       3364 NLHTLLYLGGVEPSVQLSPATNMSAHFHGCVGEVSVNGKRLDLTYSFLGSQGVGQCYDSS 3423

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              |||||||::||||||||||||||||:||||||||||||||||| ||||:||||:| ||||
Db       3424 PCERQPCRNGATCMPAGEYEFQCLCQDGFKGDLCEHEENPCQLHEPCLNGGTCRGARCLC 3483
```

```
Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||:|:|: ||||| ||||||||||||||||:|:|||  ||: ||||||||
Db       3484 LPGFSGPRCQQGAGYGVVESDWHPEGSGGNDAPGQYGAYFYDNGFLGLPGNSFSRSLPEV 3543

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||| ||||||| |||||||| | || : ||||||||||||||| |||||||||||| |
Db       3544 PETIEFEVRTSTADGLLLWQGV-VREASRSKDFISLGLQDGHLVFSYQLGSGEARLVSGD 3602

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              |||||||||:||||||:|||||||||:||:||||||||||| |   :|||||||||||| |
Db       3603 PINDGEWHRITALREGQRGSIQVDGEDLVTGRSPGPNVAVNTKDIIYIGGAPDVATLTRG 3662

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              :||||||||:||||||:||||||||||||||||||||||||||||
Db       3663 KFSSGITGCIKNLVLHTARPGAPPPQPLDLQHRAQAGANTRPCPS 3707


RESULT 3
AGRI_RAT
ID   AGRI_RAT       STANDARD;      PRT;  1959 AA.
AC   P25304; Q63034;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC   TISSUE=Embryonic spinal cord;
RX   MEDLINE=91222570; PubMed=1851019;
RA   Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
RT   "Structure and expression of a rat agrin.";
RL   Neuron 6:811-823(1991).
RN   [2]
RP   SEQUENCE OF 1777-1801 FROM N.A.
RX   MEDLINE=92407628; PubMed=1326608;
RA   Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
RT   "Structure and chromosomal localization of the mammalian agrin gene.";
RL   J. Neurosci. 12:3535-3544(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P25304-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P25304-2; Sequence=VSP_001365;
CC       Name=3;
CC         IsoId=P25304-3; Sequence=VSP_001366;
CC       Name=4;
CC         IsoId=P25304-4; Sequence=VSP_001367;
CC       Name=5;
CC         IsoId=P25304-5; Sequence=VSP_001368;
CC   -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
CC   -!- DEVELOPMENTAL STAGE: More abundant early in development.
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
```

```
CC   -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
CC       initiator.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64780; AAA40703.1; -.
DR   EMBL; M64780; AAA40702.1; ALT_INIT.
DR   EMBL; S44194; AAB23326.1; -.
DR   PIR; JH0399; AGRT.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30   1959       AGRIN.
FT   DOMAIN       65    137       KAZAL-LIKE 1.
FT   DOMAIN      141    212       KAZAL-LIKE 2.
FT   DOMAIN      213    284       KAZAL-LIKE 3.
FT   DOMAIN      287    356       KAZAL-LIKE 4.
FT   DOMAIN      361    429       KAZAL-LIKE 5.
FT   DOMAIN      430    494       KAZAL-LIKE 6.
FT   DOMAIN      495    559       KAZAL-LIKE 7.
FT   DOMAIN      563    645       KAZAL-LIKE 8.
FT   DOMAIN      688    741       LAMININ EGF-LIKE 1.
FT   DOMAIN      742    788       LAMININ EGF-LIKE 2.
FT   DOMAIN      794    864       KAZAL-LIKE 9.
FT   DOMAIN     1023   1145       SEA.
FT   DOMAIN     1220   1258       EGF-LIKE 1.
FT   DOMAIN     1263   1439       LAMININ G-LIKE 1.
FT   DOMAIN     1440   1477       EGF-LIKE 2.
FT   DOMAIN     1479   1516       EGF-LIKE 3.
FT   DOMAIN     1526   1708       LAMININ G-LIKE 2.
FT   DOMAIN     1709   1748       EGF-LIKE 4.
FT   DOMAIN     1784   1956       LAMININ G-LIKE 3.
FT   DOMAIN      869    992       SER/THR-RICH.
FT   DOMAIN     1147   1215       SER/THR-RICH.
FT   DISULFID     97    116       POTENTIAL.
FT   DISULFID    105    137       POTENTIAL.
FT   DISULFID    171    191       POTENTIAL.
FT   DISULFID    180    212       POTENTIAL.
FT   DISULFID    244    263       POTENTIAL.
FT   DISULFID    252    284       POTENTIAL.
FT   DISULFID    316    335       POTENTIAL.
FT   DISULFID    324    356       POTENTIAL.
```

```
FT   DISULFID    389    408       POTENTIAL.
FT   DISULFID    397    429       POTENTIAL.
FT   DISULFID    454    473       POTENTIAL.
FT   DISULFID    462    494       POTENTIAL.
FT   DISULFID    518    538       POTENTIAL.
FT   DISULFID    527    559       POTENTIAL.
FT   DISULFID    604    624       POTENTIAL.
FT   DISULFID    613    645       POTENTIAL.
FT   DISULFID    688    700       BY SIMILARITY.
FT   DISULFID    690    707       BY SIMILARITY.
FT   DISULFID    709    718       BY SIMILARITY.
FT   DISULFID    721    739       BY SIMILARITY.
FT   DISULFID    742    754       BY SIMILARITY.
FT   DISULFID    744    761       BY SIMILARITY.
FT   DISULFID    763    772       BY SIMILARITY.
FT   DISULFID    775    786       BY SIMILARITY.
FT   DISULFID    823    843       POTENTIAL.
FT   DISULFID    832    864       POTENTIAL.
FT   DISULFID   1224   1235       BY SIMILARITY.
FT   DISULFID   1229   1246       BY SIMILARITY.
FT   DISULFID   1248   1257       BY SIMILARITY.
FT   DISULFID   1444   1455       POTENTIAL.
FT   DISULFID   1449   1465       POTENTIAL.
FT   DISULFID   1467   1476       POTENTIAL.
FT   DISULFID   1483   1494       BY SIMILARITY.
FT   DISULFID   1488   1504       BY SIMILARITY.
FT   DISULFID   1506   1515       BY SIMILARITY.
FT   DISULFID   1713   1727       BY SIMILARITY.
FT   DISULFID   1721   1736       BY SIMILARITY.
FT   DISULFID   1738   1747       BY SIMILARITY.
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    957    957       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1144   1152       Missing (in isoform 2).
FT                                /FTId=VSP_001365.
FT   VARSPLIC   1780   1798       Missing (In isoform 3).
FT                                /FTId=VSP_001366.
FT   VARSPLIC   1788   1798       Missing (In isoform 4).
FT                                /FTId=VSP_001367.
FT   VARSPLIC   1780   1787       Missing (In isoform 5).
FT                                /FTId=VSP_001368.
FT   VARIANT     314    314       V -> VTCD (IN A VARIANT).
SQ   SEQUENCE   1959 AA;  208645 MW;  7FEFDFDAFF89CC31 CRC64;

  Query Match             22.5%;  Score 859;  DB 1;  Length 1959;
  Best Local Similarity   31.5%;  Pred. No. 1.4e-46;
  Matches  241;  Conservative   92;  Mismatches  278;  Indels  154;  Gaps   23;

Qy            2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
                : : ||    :|:|||||             |  ||::  |: || :|||| ||| | :
Db         1287 LALEFRALETEGLLLYNGNA-----------RGKDFLALALLDGRVQFRFDTGSGPAVLT 1335

Qy           62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYG-- 119
                   |:  | :| : | |   ||:| |    || | |     ||:|:  ||:|| |:
Db         1336 SLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDGLNLDTNLYVGGIPEEQVA 1395

Qy          120 -AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTA---HGISHC--PTCRDRPCQNGGQC 173
                  : :  :  |  ||:| | |  :::   |    |    |:  |    |   ||  |  |
Db         1396 MVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVGECGDHPCLPNPCHGGALC 1455

Qy          174 HDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
                   |:  ::| || |  |  |   ::  | |  |   | |     | |  | | |||||
Db         1456 QALEAGMFLCQCPPGRFGPTCADEKS-PCQPNPCHGAAPCRVLSSG-GAKCECPLGRSGT 1513

Qy          234 RCEEGVTVTTPSLSGAG-----------SYLALPALTNTHHEL----RLDVEFKPLAPDG 278
                 |:   ||    |  ||           ||| |  |     :|     |:: |    | |
Db         1514 FCQ---TV----LETAGSRPFLADFNGFSYLELKGLHTFERDLGEKMALEMVFLARGPSG 1566

Qy          279 VLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKD 338
                :||::| |:    ||||||:   |||| |:|| | ||:|| ||:||| | ||  ||  :
Db         1567 LLLYNGQKTDGKGDFVSLALHNRHLEFCYDLGKGAAVIRSKEPIALGTWVRVFLERNGRK 1626
```

```
Qy          339 GSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGE 393
                |:|:|  |  ||  || ||:      |||   ||:||      |:    :|: | | :
Db         1627 GALQVGDGPRVLGESP-KSRKVPHTMLNLKEPLYIGGAPDFSKLARGAAVSSGFSGVIQL 1685

Qy          394 VSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDL 453
                ||: | :|      |  : : :  | ||    ||
Db         1686 VSLRGHQL------LTQEHVLRAVDVSPFADHPCTQA----------------------- 1716

Qy          454 CEHEENPCQLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 509
                         |  |||:||:|   : |  |||  ||||  |::|           :| | |
Db         1717 ---------LGNPCLNGGSCVPREATYECLCPGGFSGLHCEKGL----------VEKSVG 1757

Qy          510 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV-----------PETI-------------- 544
                            |   ||| |  :   |  |           |||:
Db         1758 ------------DLETLAFDGRTYIEYLNAVIESELTNEIPAPETLDSRALFSEKALQSN 1805

Qy          545 --ELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 602
                  || :||    ||:||    :|:| :  |:::| : ||||   | |||    | |   :
Db         1806 HFELSLRTEATQGLVLW----IGKAAERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKV 1861

Qy          603 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGG 660
                |   | |: | || | ||:||  |  |:| ||     ::  |::::||     |
Db         1862 NTNRWLRIRAHREHREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPK 1921

Qy          661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                 : :|  ||::::|:          : | |   |      ||||:
Db         1922 AYGTGFVGCLRDVVVGH--------RQLHLLEDAVTKPELRPCPT 1958

RESULT 4
AGRI_CHICK
ID   AGRI_CHICK     STANDARD;      PRT;  1955 AA.
AC   P31696;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92232297; PubMed=1314620;
RA   Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahan U.J.;
RT   "cDNA that encodes active agrin.";
RL   Neuron 8:677-689(1992).
RN   [2]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=92232298; PubMed=1314621;
RA   Ruegg M.A., Tsim K.W.K., Horton S.E., Kroeger S., Escher G.,
RA   Gensch E.M., McMahan U.J.;
RT   "The agrin gene codes for a family of basal lamina proteins that
RT   differ in function and distribution.";
RL   Neuron 8:691-699(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P31696-1; Sequence=Displayed;
CC       Name=2; Synonyms=Agrin-related protein 1;
CC         IsoId=P31696-2; Sequence=VSP_001370;
CC       Name=3; Synonyms=Agrin-related protein 2;
```

```
CC         IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M94271; AAA48585.1; -.
DR   EMBL; M97371; AAA48586.1; -.
DR   EMBL; M97372; -; NOT_ANNOTATED_CDS.
DR   PIR; JH0591; AGCH.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1955       AGRIN.
FT   DOMAIN       54    126       KAZAL-LIKE 1.
FT   DOMAIN      130    201       KAZAL-LIKE 2.
FT   DOMAIN      202    273       KAZAL-LIKE 3.
FT   DOMAIN      276    344       KAZAL-LIKE 4.
FT   DOMAIN      350    418       KAZAL-LIKE 5.
FT   DOMAIN      419    483       KAZAL-LIKE 6.
FT   DOMAIN      484    548       KAZAL-LIKE 7.
FT   DOMAIN      551    633       KAZAL-LIKE 8.
FT   DOMAIN      675    728       LAMININ EGF-LIKE 1.
FT   DOMAIN      729    775       LAMININ EGF-LIKE 2.
FT   DOMAIN      781    851       KAZAL-LIKE 9.
FT   DOMAIN     1026   1148       SEA.
FT   DOMAIN     1229   1265       EGF-LIKE 1.
FT   DOMAIN     1270   1445       LAMININ G-LIKE 1.
FT   DOMAIN     1446   1483       EGF-LIKE 2.
FT   DOMAIN     1485   1522       EGF-LIKE 3.
FT   DOMAIN     1532   1713       LAMININ G-LIKE 2.
```

```
FT   DOMAIN     1714   1752       EGF-LIKE 4.
FT   DOMAIN     1776   1952       LAMININ G-LIKE 3.
FT   DOMAIN      856    995       SER/THR-RICH.
FT   DOMAIN     1150   1219       SER/THR-RICH.
FT   DISULFID     86    105       POTENTIAL.
FT   DISULFID     94    126       POTENTIAL.
FT   DISULFID    160    180       POTENTIAL.
FT   DISULFID    169    201       POTENTIAL.
FT   DISULFID    233    252       POTENTIAL.
FT   DISULFID    241    273       POTENTIAL.
FT   DISULFID    304    323       POTENTIAL.
FT   DISULFID    312    344       POTENTIAL.
FT   DISULFID    378    397       POTENTIAL.
FT   DISULFID    386    418       POTENTIAL.
FT   DISULFID    443    462       POTENTIAL.
FT   DISULFID    451    483       POTENTIAL.
FT   DISULFID    507    527       POTENTIAL.
FT   DISULFID    516    548       POTENTIAL.
FT   DISULFID    592    612       POTENTIAL.
FT   DISULFID    601    633       POTENTIAL.
FT   DISULFID    675    687       BY SIMILARITY.
FT   DISULFID    677    694       BY SIMILARITY.
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    708    726       BY SIMILARITY.
FT   DISULFID    729    741       BY SIMILARITY.
FT   DISULFID    731    748       BY SIMILARITY.
FT   DISULFID    750    759       BY SIMILARITY.
FT   DISULFID    762    773       BY SIMILARITY.
FT   DISULFID    810    830       POTENTIAL.
FT   DISULFID    819    851       POTENTIAL.
FT   DISULFID   1233   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1450   1461       BY SIMILARITY.
FT   DISULFID   1455   1471       BY SIMILARITY.
FT   DISULFID   1473   1482       BY SIMILARITY.
FT   DISULFID   1489   1500       BY SIMILARITY.
FT   DISULFID   1494   1510       BY SIMILARITY.
FT   DISULFID   1512   1521       BY SIMILARITY.
FT   DISULFID   1718   1731       BY SIMILARITY.
FT   DISULFID   1725   1740       BY SIMILARITY.
FT   DISULFID   1742   1751       BY SIMILARITY.
FT   CARBOHYD    390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1648   1651       Missing (in isoform 3).
FT                                /FTId=VSP_001369.
FT   VARSPLIC   1783   1793       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_001370.
FT   CONFLICT   1129   1131       RTI -> SIL (IN REF. 1; AAA48586).
SQ   SEQUENCE   1955 AA;  211411 MW;  B4DEB27C23422581 CRC64;

  Query Match             20.8%;  Score 796.5;  DB 1;  Length 1955;
  Best Local Similarity   30.6%;  Pred. No. 1.3e-42;
  Matches  221;  Conservative   99;  Mismatches  271;  Indels  131;  Gaps   25;

Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              | : ||     |:||||||           ||  ||||  ||||  | ||: |||   |
Db       1294 IAMEFRATELSGLLLYNGQ-----------NRGKDFISLALVGGFVELRFNTGSGTGVIT 1342

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYP-DYGA 120
                  :  | :| : : |:   | | | |   |:| |     ||:|: :|::|| | |  |
Db       1343 SKVRVEPGKWHQLVVNRNRRSGMLAV-DGEHVSGESPTGTDGLNLDTDLFVGGAPEDQMA 1401

Qy        121 I--PKAGLSSGFIGCVR---------ELRIQGEEIVFHDLNLTAHGISHC--PTCRDRPC 167
              :   :   : |  | :|         :|| :| ::::        |:  |    |   ||
Db       1402 VVAERTAATVGLKGSIRLLDVNNQMYDLREKGSDVLY------GSGVGECGNDPCHPNPC 1455

Qy        168 QNGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCH 227
               :|  ||  |:  : | |   :||  |   :   | |  |   |||:  |:| |  | |
Db       1456 HHGASCHVKEAEMFHCECLHSYTGPTCA-DERNPCDPTPCHISATCLVLPEG-GAMCACP 1513
```

```
Qy          228 LGRSGLRCEEGVTVT--------TPSLSGAGSYLALPALTN---THHELRLDVEFKPLAP 276
                :|| |  ||    ||         |  :|  ||| |  |     |  ::  ::| |   :|
Db         1514 MGREGEFCER---VTEQDHTMPFLPEFNGF-SYLELNGLQTLFLTCRQMSMEVVFLAKSP 1569

Qy          277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLN 336
                 |:: ::| |:    ||||| :  |:||:||:|| | ||||| ||:.|  |  |  ||
Db         1570 SGMIFYNGQKTDGKGDFVSLALHDGYLEYRYDLGKGAAVLRSKEPVPLNTWISVLLERSG 1629

Qy          337 KDGSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCV 391
                : | :|:| |  |:  || ||:      |||    |:||      |: |  :|  | | |
Db         1630 RKGVMRINNGERVMGESP-KSRKVPHAFLNLKEPFYVGGAPDFSKLARAAAISTSFYGAV 1688

Qy          392 GEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKG 451
                  :|: |  |      |  | |    : |                            |:
Db         1689 QRISIKGVPL------LKEQHIRSAVEIST---------------------------FRA 1715

Qy          452 DLCEHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQ-------GSGHGIA-- 498
                  |  : ||||      :||||    :   | |  ||||  |::       |    ||
Db         1716 HPCTQKPNPCQ------NGGTCSPRLESYECACQRGFSGAHCEKVIIEKAAGDAEAIAFD 1769

Qy          499 ---ESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 555
                     ::|   :   ||             | :|     ::|       ||  ::|    |
Db         1770 GRTYMEYHNAVTKSPDA-------------LDYPAEPSEKALQS--NHFELSIKTEATQG 1814

Qy          556 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 615
                |:||     |:  :  |:|:| : ||  :   | |||    | |  |||   |  : | |
Db         1815 LILWS----GKGLERSDYIALAIVDGFVQMMYDLGSKPVVLRSTVPINTNHWTHIKAYRV 1870

Qy          616 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNL 673
                 | ||:||  |  ::|  ||     ::  |::::||     ||      :|:|  ||::::
Db         1871 QREGSLQVGNEAPITGSSPLGATQLDTDGALWLGGMERLSVAHKLPKAYSTGFIGCIRDV 1930

Qy          674 VL 675
                ::
Db         1931 IV 1932
```

```
RESULT 5
AGRI_DISOM
ID   AGRI_DISOM     STANDARD;      PRT;  1328 AA.
AC   Q90404;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin (Fragment).
GN   AGRN.
OS   Discopyge ommata (Electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX   NCBI_TaxID=7785;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
RA   Snow P., McMahan U.J.;
RT   "Isolation and characterization of a cDNA that encodes an agrin
RT   homolog in the marine ray.";
RL   Mol. Cell. Neurosci. 3:406-417(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction (By similarity).
CC   -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains at least 3 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L01423; AAA49224.1; -.
DR   PIR; T43060; T43060.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 2.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 2.
DR   SMART; SM00280; KAZAL; 2.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT   NON_TER       1      1
FT   DOMAIN       79    132       LAMININ EGF-LIKE 1.
FT   DOMAIN      133    179       LAMININ EGF-LIKE 2.
FT   DOMAIN      411    533       SEA.
FT   DOMAIN      608    644       EGF-LIKE 1.
FT   DOMAIN      649    825       LAMININ G-LIKE 1.
FT   DOMAIN      865    902       EGF-LIKE 2.
FT   DOMAIN      914   1096       LAMININ G-LIKE 2.
FT   DOMAIN     1097   1135       EGF-LIKE 3.
FT   DOMAIN     1146   1324       LAMININ G-LIKE 3.
FT   DISULFID     79     91       BY SIMILARITY.
FT   DISULFID     81     98       BY SIMILARITY.
FT   DISULFID    100    109       BY SIMILARITY.
FT   DISULFID    112    130       BY SIMILARITY.
FT   DISULFID    133    145       BY SIMILARITY.
FT   DISULFID    135    152       BY SIMILARITY.
FT   DISULFID    154    163       BY SIMILARITY.
FT   DISULFID    166    177       BY SIMILARITY.
FT   DISULFID    612    623       BY SIMILARITY.
FT   DISULFID    617    632       BY SIMILARITY.
FT   DISULFID    634    643       BY SIMILARITY.
FT   DISULFID    869    880       BY SIMILARITY.
FT   DISULFID    874    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID   1101   1114       BY SIMILARITY.
FT   DISULFID   1108   1123       BY SIMILARITY.
FT   DISULFID   1125   1134       BY SIMILARITY.
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1328 AA;  144018 MW;  79D81C1AF2A71C18 CRC64;

  Query Match             17.4%;  Score 667;  DB 1;  Length 1328;
  Best Local Similarity   27.0%;  Pred. No. 1.2e-34;
  Matches  200;  Conservative  112;  Mismatches  304;  Indels  126;  Gaps   24;
```

```
Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              | : ||  : ||: |    ::  |    |   :|         |  | ||: |     |
Db        673 ISMEFRASNLDGLPLVQWTEKGKG----LHFYRPS-------EGYVELRFNHGVWDGVIT 721

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYG-- 119
                | :  |::| |   |:   | | |     : | |     ||:|: :|:||| |:
Db        722 SKTLIKPGNWHHVVGNRNRRSGMLSVDGEPHLIGESPPGTDGLNLDTDLFLGGTPEDEMT 781

Qy        120 -AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN------LTAHGISHCPT--CRDRPCQNG 170
                  :   : |  ||:| | :     : :||       |   |:  |    |   ||:|
Db        782 LVTERTTATKGLQGCIRLLDVNN---LIYDLQERSNDVLYGSGVGECGNNPCSPNPCKNR 838

Qy        171 GQCHDSESSSYVCVCPAGFTGSRC--EHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
              |:||  |:  : |     |:|  |  :|:    | |  |   | |:  |:| |  | | :
Db        839 GKCHMKEAEMFHCESVGEFSGPTCADKHNP---CDPNPCHQSANCMVLPEG-GSKCECPM 894

Qy        229 GRSGLRCE-------EGVTVTTPSLSGAGSYLALPAL----TNTHHELRLDVEFKPLAPD 277
              || |  ||       :      |  :|  ||| :  :    ::   :| ::| |    |:
Db        895 GREGELCERVSEAEQDQGKAFIPEFNGL-SYLEMNGIHTFVSDLLQKLSMEVIFLAKDPN 953

Qy        278 GVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNK 337
              |:: ::| |:    ||||| :  |:|||:|:|| | |||||  |: |  |: |: ||  :
Db        954 GMIFYNGQKTDGRGDFVSLNLRDGYLEFKYDLGKGAAVLRSKAPIPLNVWNVVTVERNGR 1013

Qy        338 DGSLRVN------GGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCV 391
               | :::|      |  |  | :|     |||    |:||       : |  : : | | :
Db       1014 KGLMKINKDELVSGESPKSRKAP--HTALNLKEAFYVGGAPDFNKFARAAGIISGFTGAI 1071

Qy        392 GEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKG 451
               ::|:       :   |  : |    : |                            |:
Db       1072 QKLSLK------SIPLLKKENIRNAMEIS---------------------------NFRW 1098

Qy        452 DLCEHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGS 507
                |    ||||      :|| |    :   |:|  |||||:|::            ||
Db       1099 HACTKTRNPCQ------NGGVCSPRLREYDCMCQRGFSGPQCEKA-----------LE-- 1139

Qy        508 GGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEA 567
                  | |     |:   |: :   |            |: ::|    ||:||     |:
Db       1140 -EKSASGSESVAFNGRTFIEYHNTVTRSEKAVQVNYFEMSIKTEATKGLILWS----GKI 1194

Qy        568 GQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEE 627
               :  |:|:| : || :   | |||    | |  |:|  :| |: | |    |::||  |
Db       1195 AEKSDYIALAVVDGFVQMTYDLGSKPVTLRSTIPVNTNQWVRIKANRIHGYGTLQVGNEA 1254

Qy        628 LVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR----FSSGITGCVKNLVLHSARPGAP 683
               |:| ||     ::  |::::||   :|   | |    :|:|  ||:|::|:
Db       1255 PVTGSSPFAATQLDTDGALWLGGIEKLA--PGNRLPKAYSTGFIGCIKDVVIDR------ 1306

Qy        684 PPQPLDLQHRAQAGANTRPCPS 705
                | | |   |        ||:
Db       1307 --QELQLVEDALNNPTILHCPA 1326
```

```
RESULT 6
UN52_CAEEL
ID   UN52_CAEEL     STANDARD;      PRT;  3375 AA.
AC   Q06561; O18261; O18263; Q9XTD2; Q9XTI5;
DT   01-JUN-1994 (Rel. 29, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane proteoglycan precursor (Perlecan homolog)
DE   (Uncoordinated protein 52).
GN   UNC-52 OR ZC101.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.
RX   MEDLINE=93339574; PubMed=8393416;
RA   Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
RT   "Products of the unc-52 gene in Caenorhabditis elegans are homologous
RT   to the core protein of the mammalian basement membrane heparan
RT   sulfate proteoglycan.";
RL   Genes Dev. 7:1471-1484(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Percy C.M., Baynes C.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RA   Durbin R.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probable role in myofilament assembly and/or attachment
CC       of the myofilament lattice to the cell membrane. May be an
CC       extracellular anchor for integrin receptors in muscle.
CC   -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=e;
CC         IsoId=Q06561-1; Sequence=Displayed;
CC         Note=No experimental confirmation available;
CC       Name=a;
CC         IsoId=Q06561-2; Sequence=VSP_007195, VSP_007196;
CC       Name=b;
CC         IsoId=Q06561-3; Sequence=VSP_007191, VSP_007192;
CC         Note=No experimental confirmation available;
CC       Name=c;
CC         IsoId=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
CC                                  VSP_007196;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Found in the basement membrane of all
CC       contractile tissues. It is concentrated over muscle dense bodies
CC       and M-lines which are associated with beta-integrin.
CC   -!- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
CC   -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 7 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L13458; AAA28156.1; -.
DR   EMBL; Z93375; CAB07567.1; -.
DR   EMBL; Z93395; CAB07567.1; JOINED.
DR   EMBL; Z93375; CAB07568.1; -.
DR   EMBL; Z93395; CAB07568.1; JOINED.
DR   EMBL; Z93375; CAB07569.1; -.
DR   EMBL; Z93395; CAB07569.1; JOINED.
DR   EMBL; Z93395; CAB07704.1; -.
DR   EMBL; Z93395; CAB07706.1; -.
DR   EMBL; Z93375; CAB07706.1; JOINED.
DR   EMBL; Z93395; CAB07707.1; -.
DR   EMBL; Z93375; CAB07707.1; JOINED.
DR   EMBL; Z93395; CAB07708.1; -.
DR   EMBL; Z93375; CAB07708.1; JOINED.
DR   HSSP; P01130; 1LDR.
DR   WormPep; ZC101.2a; CE15028.
DR   WormPep; ZC101.2b; CE15030.
DR   WormPep; ZC101.2c; CE15034.
DR   WormPep; ZC101.2e; CE18424.
DR   GO; GO:0005578; C:extracellular matrix; IEP.
DR   GO; GO:0030239; P:myofibril assembly; IEP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
```

```
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   Pfam; PF00047; ig; 16.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 5.
DR   Pfam; PF00057; ldl_recept_a; 3.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 2.
DR   SMART; SM00181; EGF; 6.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00409; IG; 17.
DR   SMART; SM00408; IGc2; 17.
DR   SMART; SM00281; LamB; 1.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50835; IG_LIKE; 17.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS01209; LDLRA_1; 3.
DR   PROSITE; PS50068; LDLRA_2; 3.
KW   Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
KW   Basement membrane; Extracellular matrix; Alternative splicing;
KW   Laminin EGF-like domain.
FT   SIGNAL          1     22       POTENTIAL.
FT   CHAIN          23   3375       BASEMENT MEMBRANE PROTEOGLYCAN.
FT   DOMAIN         45    130       IG-LIKE C2-TYPE 1.
FT   DOMAIN        148    184       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN        189    225       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN        232    269       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN        271    355       IG-LIKE C2-TYPE 2.
FT   DOMAIN        384    431       LAMININ EGF-LIKE 1 (INCOMPLETE).
FT   DOMAIN        432    441       LAMININ EGF-LIKE 2 (N-TERMINAL).
FT   DOMAIN        442    633       LAMININ DOMAIN IV 1.
FT   DOMAIN        634    666       LAMININ EGF-LIKE 2 (C-TERMINAL).
FT   DOMAIN        674    720       LAMININ EGF-LIKE 3 (INCOMPLETE).
FT   DOMAIN        721    730       LAMININ EGF-LIKE 4 (N-TERMINAL).
FT   DOMAIN        731    921       LAMININ DOMAIN IV 2.
FT   DOMAIN        922    954       LAMININ EGF-LIKE 4 (C-TERMINAL).
FT   DOMAIN        955   1004       LAMININ EGF-LIKE 5.
FT   DOMAIN       1011   1060       LAMININ EGF-LIKE 6.
FT   DOMAIN       1061   1111       LAMININ EGF-LIKE 7.
FT   DOMAIN       1126   1222       IG-LIKE C2-TYPE 3.
FT   DOMAIN       1226   1311       IG-LIKE C2-TYPE 4.
FT   DOMAIN       1319   1401       IG-LIKE C2-TYPE 5.
FT   DOMAIN       1410   1499       IG-LIKE C2-TYPE 6.
FT   DOMAIN       1503   1585       IG-LIKE C2-TYPE 7.
FT   DOMAIN       1588   1680       IG-LIKE C2-TYPE 8.
FT   DOMAIN       1690   1785       IG-LIKE C2-TYPE 9.
FT   DOMAIN       1793   1878       IG-LIKE C2-TYPE 10.
FT   DOMAIN       1886   1970       IG-LIKE C2-TYPE 11.
FT   DOMAIN       1973   2069       IG-LIKE C2-TYPE 12.
FT   DOMAIN       2073   2163       IG-LIKE C2-TYPE 13.
FT   DOMAIN       2173   2260       IG-LIKE C2-TYPE 14.
FT   DOMAIN       2263   2343       IG-LIKE C2-TYPE 15.
FT   DOMAIN       2349   2435       IG-LIKE C2-TYPE 16.
FT   DOMAIN       2446   2530       IG-LIKE C2-TYPE 17.
FT   DOMAIN       2532   2713       LAMININ G-LIKE 1.
FT   DOMAIN       2793   2960       LAMININ G-LIKE 2.
FT   DOMAIN       2961   3093       GLU-RICH.
FT   DOMAIN       2972   3066       THR-RICH.
FT   DOMAIN       3180   3359       LAMININ G-LIKE 3.
FT   DISULFID       66    114       BY SIMILARITY.
FT   DISULFID      149    161       BY SIMILARITY.
FT   DISULFID      156    174       BY SIMILARITY.
FT   DISULFID      168    183       BY SIMILARITY.
FT   DISULFID      190    202       BY SIMILARITY.
FT   DISULFID      197    215       BY SIMILARITY.
FT   DISULFID      209    224       BY SIMILARITY.
FT   DISULFID      233    246       BY SIMILARITY.
FT   DISULFID      240    259       BY SIMILARITY.
FT   DISULFID      253    268       BY SIMILARITY.
FT   DISULFID      955    964       BY SIMILARITY.
FT   DISULFID      957    971       BY SIMILARITY.
FT   DISULFID      974    983       BY SIMILARITY.
FT   DISULFID      986   1002       BY SIMILARITY.
FT   DISULFID     1011   1021       BY SIMILARITY.
FT   DISULFID     1013   1027       BY SIMILARITY.
FT   DISULFID     1030   1039       BY SIMILARITY.
FT   DISULFID     1042   1058       BY SIMILARITY.
FT   DISULFID     1061   1069       BY SIMILARITY.
FT   DISULFID     1063   1079       BY SIMILARITY.
FT   DISULFID     1082   1091       BY SIMILARITY.
FT   DISULFID     1094   1109       BY SIMILARITY.
FT   DISULFID     1152   1200       BY SIMILARITY.
FT   DISULFID     1338   1384       BY SIMILARITY.
FT   DISULFID     1435   1481       BY SIMILARITY.
FT   DISULFID     1527   1573       BY SIMILARITY.
FT   DISULFID     1618   1663       BY SIMILARITY.
FT   DISULFID     1719   1767       BY SIMILARITY.
FT   DISULFID     1814   1861       BY SIMILARITY.
FT   DISULFID     1907   1954       BY SIMILARITY.
FT   DISULFID     1998   2053       BY SIMILARITY.
FT   DISULFID     2099   2147       BY SIMILARITY.
FT   DISULFID     2195   2242       BY SIMILARITY.
FT   DISULFID     2284   2329       BY SIMILARITY.
FT   DISULFID     2374   2420       BY SIMILARITY.
FT   DISULFID     2467   2514       BY SIMILARITY.
FT   CARBOHYD     1422   1422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     2476   2476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     2950   2950       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     3143   3143       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match             13.4%;  Score 511;  DB 1;  Length 3375;
  Best Local Similarity   20.9%;  Pred. No. 2.6e-24;
  Matches  179;  Conservative  140;  Mismatches  332;  Indels  204;  Gaps   29;

Qy          1 EIKITFRPDSADGMLLYNGQKRVP-----GSPTNLANRQPDFISFGLVGGRPEFRFDAGS 55
              :|:||    :  |::     ||:       :| :  : :     :| |     : || |:
Db       2557 DIEITLNTANPKGIIFET--KRINSGDLLATPYDTIHHEAKITDYGTV----LYEFDIGN 2610

Qy         56 GMATIRHPTPLALGHFHTVTLLRSLTQGSLIVGDLA--------PVNGTSQGKFQGLDLN 107
              |   :    |:    :: : :     | :: : | :        |:   | |      :|
Db       2611 GRQIVETTNPINPNEWNVIKIKNDKNQVTIQLNDESATIRQHTNPLPSLSTG------VN 2664

Qy        108 EELYLGGYPDYGAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPC 167
                :::||  :         :: | | : :: : | :   |  : :  : :        |
Db       2665 RPVFIGGRHE-----PTNEANDFRGIISQVVLSGHNVGLGDARIPSSVVKYDACASTNLC 2719

Qy        168 QNGGQCHDSES-SSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRC 226
               ||  | :: :   : | |   | |  |:  ::  || |:|     |::  :   : | |
Db       2720 LNGANCRNANNHHGFSCECAEEFHGEYCQW-RSNSCHDESCN-TGICLDNEE--SWQCVC 2775

Qy        227 HLGRSGLRCEEGVTVTTP-SLSGAGSYLALPALTNTHHELRLDVEFKPLAPDG---VLLF 282
               || :||||||   :  |   :   |:||:           : :: :| |      :: |
Db       2776 PLGTTGLRCEEKTEIPQPLGFTSDTSFLAVKRPVKFE---SIKMKLRPQADSDEHILMYF 2832

Qy        283 SGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLR 342
              :       : : ||:::   :         | | |  : :   |   : |  |  | ::|
Db       2833 ASDYGSNTKQYTSLSLIANQVV--------LTVRRPDKEVQKIRSETLEAGEL-IDVAVR 2883

Qy        343 VNGGRPVLRSSPGKSQGLNLHTL-----LYLGGVEPSVPLSPATNMSAHFRGCVGEVSVN 397
                |   |:     :   :   ||     :::||: | :  ||   :   |:||| |: :|
Db       2884 QAGNALVMTVDGNQVSTIETDTLKPGTEIFIGGLPPGLN-SPDDVVEQSFQGCVYEILIN 2942

Qy        398 GKRLDLTYSFLGSQG-IGQCYDS-----------------------------------SPC 422
               : :||    | | | |  | :|                                   |
Db       2943 SQDVDL--QNLSSSGDISSCEESQFPVEEDDTTTTTTEEPEAVIEEPTTEEPTTTEEPI 3000
```

```
Qy        423 ERQPCQHGATC----------------------MPAGEYEFQCLCRDG------------ 448
                :| :   |                        |   ||     ||
Db       3001 TEEPTEEPTTTEEPTTTEEPTTTTEEPTTTTTEEPYHIYE---TSRDDDPEIIIPVETTT 3057

Qy        449 -------------------------FKGDLCEHEEN------------------------ 459
                                        : |: : ||
Db       3058 TSTTTTSTTEEPEAEPALVLPTDPVEENDVSDEEEEISTISTVSPDNGLDSDSDYSEGTL 3117

Qy        460 -PCQLREPCLHG---------GTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEG-SG 508
               |    |  : |           |  : |    | :|    :   |   |   : :| :
Db       3118 PPDSSSEEIVVGDVYSTQEPNNICANSTC----GMNGQCVPRNMTHYTCECKLYYDGPTC 3173

Qy        509 GNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAG 568
                  | :: | |  | |:      |     |  | :  : :|   :|:||||| :
Db       3174 SLFKPIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-QRPTVQ 3232

Qy        569 QGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEEL 628
              | :|:||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::|
Db       3233 QMEDYISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRIDNYRE 3292

Qy        629 VSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPL 688
              | ||| |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :        |
Db       3293 VDGRSTGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK--------L 3344

Qy        689 DLQHRAQAGANTRPC 703
              ||   |  | | :||
Db       3345 DLMATAIDGKNVKPC 3359

RESULT 7
CLR3_RAT
ID   CLR3_RAT       STANDARD;      PRT;  3313 AA.
AC   O88278;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Multiple
DE   epidermal growth factor-like domains 2).
GN   CELSR3 OR MEGF2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98360089; PubMed=9693030;
RA   Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
CC   -!- FUNCTION: Receptor that may have an important role in cell/cell
CC       signaling during nervous system formation.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in the brain. Expressed in
CC       cerebrellum, olfactory bulb, cerebral cortex, hippocampus and
CC       brain stem.
CC   -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC   -!- SIMILARITY: Contains 9 cadherin domains.
CC   -!- SIMILARITY: Contains 8 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 laminin EGF-like domain.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB011528; BAA32459.1; -.
```

```
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 9.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00112; CA; 9.
DR   SMART; SM00181; EGF; 6.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00232; CADHERIN_1; 7.
DR   PROSITE; PS50268; CADHERIN_2; 8.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW   Developmental protein; Hydroxylation; Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32   3313       CADHERIN EGF LAG SEVEN-PASS G-TYPE
FT                                RECEPTOR 3.
FT   DOMAIN       32   2538       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2539   2559       1 (POTENTIAL).
FT   DOMAIN     2560   2570       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2571   2591       2 (POTENTIAL).
FT   DOMAIN     2592   2599       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2600   2620       3 (POTENTIAL).
FT   DOMAIN     2621   2641       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2642   2662       4 (POTENTIAL).
FT   DOMAIN     2663   2679       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2680   2700       5 (POTENTIAL).
FT   DOMAIN     2701   2724       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2725   2745       6 (POTENTIAL).
FT   DOMAIN     2746   2752       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2753   2773       7 (POTENTIAL).
FT   DOMAIN     2774   3313       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      317    424       CADHERIN 1.
FT   DOMAIN      425    536       CADHERIN 2.
FT   DOMAIN      537    642       CADHERIN 3.
FT   DOMAIN      643    747       CADHERIN 4.
FT   DOMAIN      748    849       CADHERIN 5.
FT   DOMAIN      850    952       CADHERIN 6.
FT   DOMAIN      953   1058       CADHERIN 7.
FT   DOMAIN     1059   1160       CADHERIN 8.
FT   DOMAIN     1161   1257       CADHERIN 9.
FT   DOMAIN     1366   1424       EGF-LIKE 1, CALCIUM-BINDING.
FT   DOMAIN     1426   1462       EGF-LIKE 2, CALCIUM-BINDING.
FT   DOMAIN     1466   1505       EGF-LIKE 3, CALCIUM-BINDING.
```

```
FT   DOMAIN      1506   1710       LAMININ G-LIKE 1.
FT   DOMAIN      1713   1749       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      1753   1935       LAMININ G-LIKE 2.
FT   DOMAIN      1937   1972       EGF-LIKE 5, CALCIUM-BINDING.
FT   DOMAIN      1973   2011       EGF-LIKE 6, CALCIUM-BINDING.
FT   DOMAIN      2012   2044       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN      2046   2081       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN      2087   2120       LAMININ EGF-LIKE.
FT   DOMAIN      2475   2527       GPS.
FT   DISULFID    1370   1381       BY SIMILARITY.
FT   DISULFID    1375   1412       BY SIMILARITY.
FT   DISULFID    1414   1423       BY SIMILARITY.
FT   DISULFID    1430   1441       BY SIMILARITY.
FT   DISULFID    1435   1450       BY SIMILARITY.
FT   DISULFID    1452   1461       BY SIMILARITY.
FT   DISULFID    1470   1481       BY SIMILARITY.
FT   DISULFID    1475   1491       BY SIMILARITY.
FT   DISULFID    1493   1504       BY SIMILARITY.
FT   DISULFID    1717   1728       BY SIMILARITY.
FT   DISULFID    1722   1737       BY SIMILARITY.
FT   DISULFID    1739   1748       BY SIMILARITY.
FT   DISULFID    1941   1952       BY SIMILARITY.
FT   DISULFID    1946   1961       BY SIMILARITY.
FT   DISULFID    1963   1972       BY SIMILARITY.
FT   DISULFID    1976   1987       BY SIMILARITY.
FT   DISULFID    1981   1999       BY SIMILARITY.
FT   DISULFID    2001   2010       BY SIMILARITY.
FT   DISULFID    2018   2031       BY SIMILARITY.
FT   DISULFID    2033   2043       BY SIMILARITY.
FT   DISULFID    2050   2065       BY SIMILARITY.
FT   DISULFID    2052   2068       BY SIMILARITY.
FT   DISULFID    2070   2080       BY SIMILARITY.
FT   MOD_RES     1954   1954       HYDROXYLATION (POTENTIAL).
FT   CARBOHYD     623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     838    838       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1173   1173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1213   1213       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1308   1308       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1318   1318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1640   1640       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1704   1704       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1761   1761       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2044   2044       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2173   2173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2192   2192       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2382   2382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2472   2472       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2504   2504       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE    3313 AA;  359348 MW;  B11DA09517288764 CRC64;

  Query Match             11.5%;  Score 441;  DB 1;  Length 3313;
  Best Local Similarity   23.9%;  Pred. No. 6.8e-20;
  Matches  187;  Conservative   87;  Mismatches  256;  Indels  252;  Gaps   42;

Qy          61 RHPTPLALGHF--HTVTLLRSLTQGSLI--------VGD---------LAPVNGTSQGK- 100
               |  :|| ||||      :| : |:   |        ||          |||  |   |
Db        1277 RFLSPL-LGHFLEGVAAVLATPTEDVFIFNIQNDTDVGGTVLNVSFSALAP-RGAGAGAA 1334

Qy         101 ---FQGLDLNEELYLGGYPDYGAIPKAGLSS---------------------GFIGCVREL 137
                  |   :| |:||         : :| |::                     :: ||  |
Db        1335 GPWFSSEELQEQLY---------VRRAALAARSLLDVLPFDDNVCLREPCENYMKCVSVL 1385

Qy         138 R-------IQGEEIVFHDLNLTAHGISHCP-------------TCRDRPCQNGGQCHDSE 177
               |       :     :|  :   |     ||              |   ||:|||  |   |
Db        1386 RFDSSAPFLASASTLFRPIQPIAGLRCRCPPGFTGDFCETELDLCYSNPCRNGGACARRE 1445

Qy         178 SSSYVCVCPAGFTGSRCE-HSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRS--GLR 234
                  | |||   |||  ||  ::|  | |  |    || | |:| |: |:|  | :  | |
Db        1446 -GGYTCVCRPRFTGEDCELDTEAGRCVPGVCRNGGTCTNAPNG-GFRCQCPAGGAFEGPR 1503

Qy         235 CEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFV 294
               ||       ||     |::    |    | | | : |  : | | |||  |:     ||:
Db        1504 CEVAARSFPPS-----SFVMFRGLRQRFH-LTLSLSFATVQPSG-LLFYNGRLNEKHDFL 1556
```

```
Qy         295 SLAMVGGHLEFRYELGSGLAVLRSAEP--LALGRWHRVSAERLNKDGSLRVNGGRPVLRS 352
               :| :| | :   |  |    |:    |  |: |:|| |     ||  :  : | :
Db        1557 ALELVAGQVRLTYSTGESSTVVSPTVPGGLSDGQWHTVHLRYYNKPRTDALGGAQ----- 1611

Qy         353 SPGK--------------------------------------SQGLNLHTLLYLGGVEPSV 375
                | |                                      : |:|   | |||| |::
Db        1612 GPSKDKVAVLSVDDCNVAVALRFGAEIGNYSCAAAGVQTSSKKSLDLTGPLLLGGV-PNL 1670

Qy         376 PLS-PATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCM 434
               | : | :     | ||: :: ::|:|:|:  :|: : |                  | |
Db        1671 PENFPVSRKD--FIGCMRDLHIDGRRVDMA-AFVANNGT---------------TAGCQ 1711

Qy         435 PAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ----GTRCLCLPGFSGPRCQ 490
                   :     |  |                 || :|| |     |  | |  || |  |:
Db        1712 AKSHF-----CASG-----------------PCKNGGLCSERWGGFSCDCPVGFGGKDCR 1749

Qy         491 QGSGHGIAESDWHLEGSG------GNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETI 544
                   |      :| :|:|      ||| |                          ||  :
Db        1750 LTMAH-----PYHFQGNGTLSWDFGNDMP----------------------VSVPWYL 1780

Qy         545 ELEVRTSTASGLLLWQGVEVG-------EAGQGKDFISLGLQDGHLVFRYQLGSGEARLV 597
                |  ||    |:|:   |::|       :  ||   ::|    || |           |:
Db        1781 GLSFRTRATKGVLM--QVQLGPHSVLLCKLDQGLLSVTLSRASGHAVH---------LLL 1829

Qy         598 SEDPINDGEWH--RVTALRE--GRRG----SIQVDGEELVSGRSPGPNVAVNAKGSVYIG 649
                :  ::|| ||  |:    |  ||||     : :|        : |  :       :::|
Db        1830 DQMTVSDGRWHDLRLELQEEPGGRRGHHIFMVSLDFTLFQDTMAMGSELEGLKVKHLHVG 1889

Qy         650 GAPDVATLTGGRFSSGITGCVKNLVLHSARPGA---PPPQPLDLQHR--AQAGAN-TRPC 703
               | |  :   |     |: ||:: :       |:   |||      ||   : |   | ||
Db        1890 GPPPSSKEEG---PQGLVGCIQGVWTGFTPFGSSALPPP-----SHRINVEPGCTVTNPC 1941

Qy         704 PS 705
               |
Db        1942 AS 1943

RESULT 8
NX2A_RAT
ID   NX2A_RAT      STANDARD;      PRT;  1715 AA.
AC   Q63374; Q63375;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 2-alpha precursor (Neurexin II-alpha).
GN   NRXN2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., VARIANT LEU-434, AND ALTERNATIVE SPLICING.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=92320296; PubMed=1621094;
RA   Ushkaryov Y.A., Petrenko A.G., Geppert M., Suedhof T.C.;
RT   "Neurexins: synaptic cell surface proteins related to the alpha-
RT   latrotoxin receptor and laminin.";
RL   Science 257:50-56(1992).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=95209856; PubMed=7695896;
RA   Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT   "Cartography of neurexins: more than 1000 isoforms generated by
RT   alternative splicing and expressed in distinct subsets of neurons.";
RL   Neuron 14:497-507(1995).
RN   [3]
RP   INTERACTION WITH NEUREXOPHILIN 1.
RX   MEDLINE=99074239; PubMed=9856994;
RA   Missler M., Hammer R.E., Suedhof T.C.;
RT   "Neurexophilin binding to alpha-neurexins. A single LNS domain
RT   functions as an independently folding ligand-binding unit.";
```

```
RL   J. Biol. Chem. 273:34716-34723(1998).
RN   [4]
RP   INTERACTION WITH ALPHA-DYSTROGLYCAN.
RX   MEDLINE=21363578; PubMed=11470830;
RA   Sugita S., Saito F., Tang J., Satz J., Campbell K., Suedhof T.C.;
RT   "A stoichiometric complex of neurexins and dystroglycan in brain.";
RL   J. Cell Biol. 154:435-445(2001).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: The laminin G-like domain 1 binds to neurexophilin 1.
CC       Isoforms alpha 2C bind to alpha-dystroglycan.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms, alpha-type (shown here) and
CC         beta-type, are produced by use of alternative promoters.
CC         Beta-type isoforms differ from alpha-type isoforms in their
CC         N-terminus;
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=At least 216 isoforms may be produced by alternative
CC         splicing. There is a combination of five alternatively spliced
CC         extracellular domains at sites 1 to 5, each consisting of
CC         modular sequences (A-C) that seem to be used independently.
CC         Additional isoforms may derive from a minor cytoplasmic splice -
CC         site 6. Beta isoforms (AC Q63376) share the combination of
CC         alternatively spliced domains at sites 4, 5 and 6;
CC       Name=Alpha 1A2A3A4A5A;
CC         IsoId=Q63374-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Brain (neuronal synapse).
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M96376; AAA41706.1; -.
DR   EMBL; M96376; AAA41707.1; -.
DR   PIR; C40228; C40228.
DR   HSSP; Q63373; 1C4R.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 5.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 6.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage;
KW   Polymorphism.
FT   SIGNAL        1     29       BY SIMILARITY.
FT   CHAIN        30   1715       NEUREXIN 2-ALPHA.
FT   DOMAIN       30   1639       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1640   1660       POTENTIAL.
FT   DOMAIN     1661   1715       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       30    206       LAMININ G-LIKE 1.
FT   DOMAIN      202    242       EGF-LIKE 1.
FT   DOMAIN      289    486       LAMININ G-LIKE 2.
FT   DOMAIN      493    686       LAMININ G-LIKE 3.
FT   DOMAIN      690    727       EGF-LIKE 2.
FT   DOMAIN      732    907       LAMININ G-LIKE 4.
FT   DOMAIN      921   1096       LAMININ G-LIKE 5.
FT   DOMAIN     1099   1136       EGF-LIKE 3.
FT   DOMAIN     1140   1348       LAMININ G-LIKE 6.
FT   DOMAIN     1371   1374       POLY-THR.
FT   DOMAIN     1377   1380       POLY-THR.
FT   DOMAIN     1447   1450       POLY-PRO.
FT   DOMAIN     1647   1650       POLY-ALA.
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    295    295       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    338    338       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1239   1239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    250    259       Missing (in some isoforms Alpha 1B).
FT                                /FTId=VSP_003509.
FT   VARSPLIC    250    283       Missing (in some isoforms Alpha 1C).
FT                                /FTId=VSP_003510.
FT   VARSPLIC    385    399       Missing (in some isoforms Alpha 2C).
FT                                /FTId=VSP_003511.
FT   VARSPLIC    393    399       Missing (in some isoforms Alpha 2B).
FT                                /FTId=VSP_003512.
FT   VARSPLIC    797    806       DCLRVGCAPS -> G (in some isoforms Alpha
FT                                3C).
FT                                /FTId=VSP_003513.
FT   VARSPLIC    807    809       Missing (in some isoforms Alpha 3B).
FT                                /FTId=VSP_003514.
FT   VARSPLIC   1256   1285       Missing (in some isoforms Alpha 4B).
FT                                /FTId=VSP_003515.
FT   VARSPLIC   1421   1614       Missing (in some isoforms Alpha 5B).
FT                                /FTId=VSP_003516.
FT   VARSPLIC   1667   1715       DEGSYQVDQSRNYISNSAQSNGAVVKEKAPAAPKTPSKAKK
FT                                NKDKEYYV -> CRKSPREEKLLPGSAQGLGLDLAKACCVC
FT                                RCRATCIAGKPLEERGGGRGEGERQMQIYIKNK (in
FT                                some isoforms Alpha 6).
FT                                /FTId=VSP_003517.
FT   VARIANT     434    434       N -> L.
SQ   SEQUENCE   1715 AA;  185282 MW;  59FBF18661F3DB15 CRC64;

  Query Match             11.0%;  Score 421;  DB 1;  Length 1715;
  Best Local Similarity   21.6%;  Pred. No. 5.9e-19;
  Matches  193;  Conservative  114;  Mismatches  274;  Indels  312;  Gaps   37;

Qy          2 IKITFRPDSADGMLLYN-GQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              | : ||    :|:||:: |::   |  :: :::: |: :  |: |      | |||   :
Db        517 ISLDFRTTEPNGLLLFSQGRRAGAGVGSHSSSQRADYFAMELLDGYLYLLLDMGSGGIKL 576

Qy         61 RHPT-PLALGHFHTVTLLRSLTQGSLIVGDLA-PVNGTSQGKFQGLDLNEELYLGGYPDY 118
              |  :  :  | :  |   |   :||: |   : |   |  |: : |||  |||||| |:
Db        577 RASSRKVNDGEWCHVDFQRDGRKGSISVNSRSTPFLAT--GESEVLDLESELYLGGLPEG 634

Qy        119 GAIP--------KAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCP--------TC 162
              | :          | | :|:||||:| | |       |       :   |         |
Db        635 GRVDLPLPPEVWTAALRAGYVGCVRDLFIDGRSRDLRGLAEAQGAVGVAPFCSRETLKQC 694

Qy        163 RDRPCQNGGQCHDSESSSYVCVC-PAGFTGSRCE----------------------HSQA 199
                 ||:||| | :   : :|| |   || |  ||                      |::|
Db        695 ASAPCRNGGICREG-WNRFVCDCIGTGFLGRVCEREATVLSYDGSMYMKIMLPNAMHTEA 753

Qy        200 ------------------------------------------LHCHPEACGPDATC---- 213
                                                        | |    | | |
Db        754 EDVSLRFMSQRAYGLMMATTSRESADTLRLELDGGQMKLTVNLDCLRVGCAPSAAAKGPE 813

Qy        214 ------------------VNRP-------------------------------------- 217
                                | |
Db        814 TLFAGHKLNDNEWHTLRVVRRGKSLQLSVDNVTVEGQMAGAHTRLEFHNIETGIMTERRF 873

Qy        218 -----------------DGRGYTCRCHLG---------RSGLRCEEGVTVTTPSLSGAGS 251
                               :|: |  :|  |         | |||      ||  | |   |
Db        874 ISVVPSNFIGHLSGLVFNGQPYMDQCKDGDITYCELNARFGLRAIVADPVTFKSRS---S 930

Qy        252 YLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGS 311
              ||||  |   :  : |  :||  ||||:|||: |      ||: : :| |:: : ::||:
Db        931 YLALATL-QAYASMHLFFQFKTTAPDGLLLFNSGNG---NDFIVIELVKGYIHYVFDLGN 986
```

```
Qy      312 GLAVLR--SAEPLALGRWHRVSAER-LNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYL 368
              ::::  | :|:   :|| |   |      :|:::  | | : | | :: |:|   ||:
Db      987 SPSLMKGNSDKPVNDNQWHNVVVSRDPGNVHTLKID-SRTVTQHSNG-ARNLDLKGELYI 1044

Qy      369 GGVEPSVPLSPATNMSAH---FRGCVGEVSVNGKRLDLTYSFLGSQGIGQCY-----DSS 420
            ||:  ::  |    : |    |:||:  | :||:  ||    |    |||        |:
Db     1045 GGLSKNM-FSNLPKLVASRDGFQGCLASVDLNGRLPDLIADAL--HRIGQVERGCDGPST 1101

Qy      421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
             |  : | :   |:            |||                      ||  |
Db     1102 TCTEESCANQGVCLQQW---------DGF----------------------TCDCT---- 1126

Qy      481 LPGFSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVF 533
            :  : || |         | |  :    |       || |        |   : |  |
Db     1127 MTSYGGPVCNDPGTTYIFGKGGALITYTW-----PPNDRPSTR----MDRLAVGFSTHQR 1177

Qy      534 SRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGE 593
            |  |          ||  :||||               |:: | :  | :   : :|: :
Db     1178 SAVL----------VRVDSASGL--------------GDYLQLHIDQGTVGVIFNVGTDD 1213

Qy      594 ARLVSEDP---INDGEWHRVTALREGRRGSIQVD----GEELVSGRSPGPNVAV------ 640
              :  ::|   ::||::| |   | |   ::|||     |   :|      :|:
Db     1214 --ITIDEPNAIVSDGKYHVVRFTRSGGNATLQVDSWPVNERYPAGNFDNERLAIARQRIP 1271

Qy      641 ---------------------NAKGSVYIGGAPDVATLTGGRFSSGITGCVKN 672
                                 |:: :: |||        |  |   ::|   |
Db     1272 YRLGRVVDEWLLDKGRQLTIFNSQAAIKIGGRDQ-----GRPFQGQVSGLYYN 1319

RESULT 9
NX2A_HUMAN
ID   NX2A_HUMAN     STANDARD;      PRT;  1712 AA.
AC   Q9P2S2; Q9Y2D6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 2-alpha precursor (Neurexin II-alpha).
GN   NRXN2 OR KIAA0921.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Fetal brain;
RA   Seki N., Yoshikawa T., Azuma T., Muramatsu M., Saito T.;
RT   "Human neurexin II.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
RA   Burke J., Dors M., Fleetwood P., Kaur A., Madan A., Nesbitt R.,
RA   Pate D., Hood L.;
RT   "Sequencing of human neurexin II gene.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND REVISIONS.
RC   TISSUE=Brain;
RX   MEDLINE=22158633; PubMed=12168954;
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
RL   DNA Res. 9:99-106(2002).
RN   [4]
RP   SEQUENCE OF 272-1712 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: The laminin G-like domain 1 binds to neurexophilin 1.
CC       Specific isoforms bind to alpha-dystroglycan (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms, alpha-type (shown here) and
CC         beta-type (AC P58401), are produced by use of alternative
CC         promoters. Beta-type isoforms differ from alpha-type isoforms
CC         in their N-terminus;
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q9P2S2-1; Sequence=Displayed;
CC       Name=2; Synonyms=Alpha-2B;
CC         IsoId=Q9P2S2-2; Sequence=VSP_003505, VSP_003506, VSP_003507,
CC                                  VSP_003508;
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in brain.
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB035266; BAA94075.1; -.
DR   EMBL; AC044790; AAK68154.1; -.
DR   EMBL; AB023138; BAA76765.2; ALT_INIT.
DR   HSSP; Q63373; 1C4R.
DR   Genew; HGNC:8009; NRXN2.
DR   MIM; 600566; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 5.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 6.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   1712       NEUREXIN 2-ALPHA.
FT   DOMAIN       29   1636       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1637   1657       POTENTIAL.
FT   DOMAIN     1658   1712       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       29    206       LAMININ G-LIKE 1.
FT   DOMAIN      202    242       EGF-LIKE 1.
FT   DOMAIN      289    486       LAMININ G-LIKE 2.
FT   DOMAIN      493    686       LAMININ G-LIKE 3.
FT   DOMAIN      690    727       EGF-LIKE 2.
FT   DOMAIN      732    904       LAMININ G-LIKE 4.
FT   DOMAIN      918   1093       LAMININ G-LIKE 5.
FT   DOMAIN     1096   1133       EGF-LIKE 3.
FT   DOMAIN     1137   1345       LAMININ G-LIKE 6.
FT   DOMAIN     1368   1371       POLY-THR.
FT   DOMAIN     1374   1377       POLY-THR.
FT   DOMAIN     1444   1447       POLY-PRO.
FT   DOMAIN     1644   1647       POLY-ALA.
```

```
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    338    338       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    841    841       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1236   1236       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    260    283       Missing (in isoform 2).
FT                                /FTId=VSP_003505.
FT   VARSPLIC    393    399       Missing (in isoform 2).
FT                                /FTId=VSP_003506.
FT   VARSPLIC    797    806       DCLRVGCAPS -> G (in isoform 2).
FT                                /FTId=VSP_003507.
FT   VARSPLIC   1253   1282       Missing (in isoform 2).
FT                                /FTId=VSP_003508.
SQ   SEQUENCE   1712 AA;  184980 MW;  E4E8EC404DA1D0B0 CRC64;

  Query Match             11.0%;  Score 420.5;  DB 1;  Length 1712;
  Best Local Similarity   21.6%;  Pred. No. 6.3e-19;
  Matches  192;  Conservative  114;  Mismatches  275;  Indels  309;  Gaps   37;

Qy        2 IKITFRPDSADGMLLYN-GQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
            | : ||    :|:||:: |::   |: :: : :: |: :  |: |      | |||   :
Db      517 ISLDFRTTEPNGLLLFSQGRRAGGGAGSHSSAQRADYFAMELLDGHLYLLLDMGSGGIKL 576

Qy       61 RHPT-PLALGHFHTVTLLRSLTQGSLIVGDLA-PVNGTSQGKFQGLDLNEELYLGGYPDY 118
            |  :  :  | :  |   |   :||: |   : |   |  |  : |||  |||||| |:
Db      577 RASSRKVNDGEWCHVDFQRDGRKGSISVNSRSTPFLAT--GDSEILDLESELYLGGLPEG 634

Qy      119 GAIP--------KAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCP--------TC 162
            | :          | | :|::||||:| | |       |      :   |         |
Db      635 GRVDLPLPPEVWTAALRAGYVGCVRDLFIDGRSRDLRGLAEAQGAVGVAPFCSRETLKQC 694

Qy      163 RDRPCQNGGQCHDSESSSYVCVC-PAGFTGSRCE----------------------HSQA 199
               ||:||| | :   : ::| |   || |  ||                      |::|
Db      695 ASAPCRNGGVCREG-WNRFICDCIGTGFLGRVCEREATVLSYDGSMYMKIMLPNAMHTEA 753

Qy      200 ------------------------------------------LHCHPEACGPD------- 210
                                                      | |    | |
Db      754 EDVSLRFMSQRAYGLMMATTSRESADTLRLELDGGQMKLTVNLDCLRVGCAPSKGPETLF 813

Qy      211 ------------ATCVNRP----------------------------------------- 217
                           | |
Db      814 AGHKLNDNEWHTVRVVRRGKSLQLSVDNVTVEGQMAGAHMRLEFHNIETGIMTERRFISV 873

Qy      218 --------------DGRGYTCRCHLG---------RSGLRCEEGVTVTTPSLSGAGSYLA 254
                          :|: |  :|  |         | |||      ||  | |   ||||
Db      874 VPSNFIGHLSGLVFNGQPYMDQCKDGDITYCELNARFGLRAIVADPVTFKSRS---SYLA 930

Qy      255 LPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA 314
            |  |   :  : |  :||  ||||:|||: |      ||: : :| |:: : ::||:| :
Db      931 LATL-QAYASMHLFFQFKTTAPDGLLLFNSGNG---NDFIVIELVKGYIHYVFDLGNGPS 986

Qy      315 VLR--SAEPLALGRWHRVSAER-LNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGV 371
            :::  | :|:   :|| |   |      :|:::  | | : | | :: |:|   ||:||:
Db      987 LMKGNSDKPVNDNQWHNVVVSRDPGNVHTLKID-SRTVTQHSNG-ARNLDLKGELYIGGL 1044

Qy      372 EPSVPLSPATNMSAH---FRGCVGEVSVNGKRLDLTYSFLGSQGIGQCY-----DSSPCE 423
              ::  |    : |    |:||:  | :||:  ||    |    |||        |: |
Db     1045 SKNM-FSNLPKLVASRDGFQGCLASVDLNGRLPDLIADAL--HRIGQVERGCDGPSTTCT 1101

Qy      424 RQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPG 483
             : | :   |:            |||                      ||  |    :
Db     1102 EESCANQGVCLQQW---------DGF----------------------TCDCT----MTS 1126

Qy      484 FSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRS 536
            : || |         | |  :    |       || |        |   : |  |  |
Db     1127 YGGPVCNDPGTTYIFGKGGALITYTW-----PPNDRPSTR----MDRLAVGFSTHQRSAV 1177

Qy      537 LPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARL 596
            |          ||  :||||               |:: | :  | :   : :|: :  :
Db     1178 L----------VRVDSASGL--------------GDYLQLHIDQGTVGVIFNVGTDD--I 1211

Qy      597 VSEDP---INDGEWHRVTALREGRRGSIQVD----GEELVSGRSPGPNVAV--------- 640
              ::|   ::||::| |   | |   :|||     |   :|      :|:
Db     1212 TIDEPNAIVSDGKYHVVRFTRSGGNATLQVDSWPVNERYPAGNFDNERLAIARQRIPYRL 1271
```

```
Qy      641 ------------------NAKGSVYIGGAPDVATLTGGRFSSGITGCVKN 672
                              |:: :: |||        |  |   ::|   |
Db     1272 GRVVDEWLLDKGRQLTIFNSQAAIKIGGRDQ-----GRPFQGQVSGLYYN 1316

RESULT 10
NX1A_HUMAN
ID   NX1A_HUMAN     STANDARD;      PRT;  1477 AA.
AC   Q9ULB1; O60323; Q9C079; Q9C080; Q9C081; Q9H3M2; Q9UDM6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 1-alpha precursor (Neurexin I-alpha).
GN   NRXN1 OR KIAA0578.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Seki N., Yoshikawa T., Azuma T., Saito T., Muramatsu M.;
RT   "Human neurexin I-alpha.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 105-1477 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [3]
RP   SEQUENCE OF 441-1121 FROM N.A.
RA   Mead K., Wohldmann P., Minges B.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1083-1477 FROM N.A.
RA   Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
RA   Dors M., Dickhoff R., Fleetwood P., Harrison G., Kaur A., Madan A.,
RA   Nesbitt R., Traicoff R., Hood L.;
RT   "Sequencing of human chromosome 2 neurexin I gene.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: The cytoplasmic C-terminal region binds to CASK. Laminin
CC       G-like domain 2 binds to NXPH1. Specific isoforms bind to alpha-
CC       dystroglycan and to alpha-latroxin. Binds SYTL1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms are produced by use of alternative
CC         promoters. The alpha and beta (AC P58400) isoforms differ in
CC         their N-terminus;
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=A number of isoforms may be produced;
CC       Name=1;
CC         IsoId=Q9ULB1-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Heart and brain.
CC   -!- PTM: N- and O-glycosylated (By similarity).
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB035356; BAA87821.1; -.
DR   EMBL; AB011150; BAA25504.1; -.
DR   EMBL; AC007462; AAF03536.1; -.
DR   EMBL; AC068725; AAG59602.1; -.
DR   EMBL; AC069550; AAG38120.1; -.
DR   EMBL; AC078994; AAK06387.1; -.
DR   EMBL; AC068715; AAG59642.1; -.
DR   HSSP; Q63373; 1C4R.
DR   Genew; HGNC:8008; NRXN1.
DR   MIM; 600565; -.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 6.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 6.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage.
FT   SIGNAL        1     30       BY SIMILARITY.
FT   CHAIN        31   1477       NEUREXIN 1-ALPHA.
FT   DOMAIN       31   1401       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1402   1422       POTENTIAL.
FT   DOMAIN     1423   1477       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    217       LAMININ G-LIKE 1.
FT   DOMAIN      213    256       EGF-LIKE 1.
FT   DOMAIN      283    473       LAMININ G-LIKE 2.
FT   DOMAIN      480    672       LAMININ G-LIKE 3.
FT   DOMAIN      676    713       EGF-LIKE 2.
FT   DOMAIN      718    891       LAMININ G-LIKE 4.
FT   DOMAIN      905   1080       LAMININ G-LIKE 5.
FT   DOMAIN     1083   1120       EGF-LIKE 3.
FT   DOMAIN     1126   1294       LAMININ G-LIKE 6.
FT   DOMAIN     1324   1327       POLY-THR.
FT   DOMAIN     1409   1412       POLY-ALA.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    790    790       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1223   1223       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1477 AA;  161882 MW;  FF845FB428B1A683 CRC64;

  Query Match             10.8%;  Score 412.5;  DB 1;  Length 1477;
  Best Local Similarity   20.8%;  Pred. No. 1.7e-18;
  Matches  200;  Conservative  113;  Mismatches  287;  Indels  361;  Gaps   41;

Qy          2 IKITFRPDSADGMLLY-NGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              |   ||    :|::|: :|: |      :    : || :  :: |      | |||   |
Db        504 ISFDFRTTEPNGLILFSHGKPRHQKDAKHPQMIKVDFFAIEMLDGHLYLLLDMGSGTIKI 563

Qy         61 RHPTPLAL------GHFHTVTLLRSLTQGSLIVGDL-APVNGTSQGKFQGLDLNEELYLG 113
              :     ||      | :: |   |    |:: |  |  |   |: |: : |||::|||||
Db        564 K-----ALLKKVNDGEWYHVDFQRDGRSGTISVNTLRTPY--TAPGESEILDLDDELYLG 616

Qy        114 GYPDYGAIPKAG-----------LSSGFIGCVRELRIQGEEIVFH---DLNLTAHGISHC 159
              | |:     |||           |: |::||:|:| | |:        ::  ||     |
Db        617 GLPE----NKAGLVFPTEVWTALLNYGYVGCIRDLFIDGQSKDIRQMAEVQSTAGVKPSC 672

Qy        160 -----PTCRDRPCQNGGQCHDSESSSYVCVCP-AGFTGSRCE------------------ 195
                     |   ||:| | | |   : ||| |   |: |  ||
Db        673 SKETAKPCLSNPCKNNGMCRDG-WNRYVCDCSGTGYLGRSCEREATVLSYDGSMFMKIQL 731

Qy        196 ----HSQA------------------------------------------LHCHPEAC-- 207
                  |::|                                          | |    |
Db        732 PVVMHTEAEDVSLRFRSQRAYGILMATTSRDSADTLRLELDAGRVKLTVNLDCIRINCNS 791

Qy        208 --GPDATCV--NRPDGRGYTCR-------------------------------------- 225
                ||:      |  |   :| |
Db        792 SKGPETLFAGYNLNDNEWHTVRVVRRGKSLKLTVDDQQAMTGQMAGDHTRLEFHNIETGI 851

Qy        226 ---------------------------------------CHL-GRSGLR--CEEGVTVTT 243
                                                     | |  | | |    : ||  |
Db        852 ITERRYLSSVPSNFIGHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIIADPVTFKT 911

Qy        244 PSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHL 303
               |     ||:||  |   :  : |  :||  : ||::|::   ||   ||: : :| |:|
Db        912 KS-----SYVALATL-QAYTSMHLFFQFKTTSLDGLILYN---SGDGNDFIVVELVKGYL 962

Qy        304 EFRYELGSGLAVLR--SAEPLALGRWHRVSAER---------LNKDGSLRVNGGRPVLRS 352
               : ::||:|  :::  | :||   :|| |   |         ::   : ::  |
Db        963 HYVFDLGNGANLIKGSSNKPLNDNQWHNVMISRDTSNLHTVKIDTKITTQITAG------ 1016

Qy        353 SPGKSQGLNLHTLLYLGGV--EPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGS 410
                  :: |:| : ||:|||  |    |    :    |:||:  | :||:  ||
Db       1017 ----ARNLDLKSDLYIGGVAKETYKSLPKLVHAKEGFQGCLASVDLNGRLPDL------- 1065

Qy        411 QGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHG 470
                     |:  |              |: |         :|  ||     ||  : | :
Db       1066 -----ISDALFCN-------------GQIE---------RG--CEGPSTTCQ-EDSCSNQ 1095

Qy        471 GTC----QGTRCLC-LPGFSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGA 518
              | |     |  | | :  |||| |           | |     |       || |
Db       1096 GVCLQQWDGFSCDCSMTSFSGPLCNDPGTTYIFSKGGGQITYKW-----PPNDRPSTRA- 1149

Qy        519 YFHDDGFLAFPGHVFSRSLPEVPETIELE---VRTSTASGLLLWQGVEVGEAGQGKDFIS 575
                 |   : |              |:: |   ||  ::|||               |::
Db       1150 ---DRLAIGF-------------STVQKEAVLVRVDSSSGL--------------GDYLE 1179

Qy        576 LGLQDGHLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDGEELVSGRSP 634
              | :  | :  :: :|: :  :   :  ||||::| |   | |   ::|||   ::
Db       1180 LHIHQGKIGVKFNVGTDDIAIEESNAIINDGKYHVVRFTRSGGNATLQVDSWPVIERYPA 1239

Qy        635 GPNVAV-NAKGSVYIGG------------------------------------------- 650
              |  : : |:: :: |||
Db       1240 GRQLTIFNSQATIIIGGKEQGQPFQGQLSGLYYNGLKVLNMAAENDANIAIVGNVRLVGE 1299

Qy        651 ------APDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 704
                        ||      |: |      |   :|| | || :    |        :  ||
Db       1300 VPSSMTTESTATAMQSEMSTSIMETTTTLATSTARRGKPPTKEPISQTTDDILVASAECP 1359

Qy        705 S 705
              |
Db       1360 S 1360

RESULT 11
NX1A_BOVIN
ID   NX1A_BOVIN     STANDARD;      PRT;  1530 AA.
AC   Q28146;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 1-alpha precursor (Neurexin I-alpha).
GN   NRXN1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=95209856; PubMed=7695896;
RA   Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT   "Cartography of neurexins: more than 1000 isoforms generated by
```

```
RT   alternative splicing and expressed in distinct subsets of neurons.";
RL   Neuron 14:497-507(1995).
RN   [2]
RP   INTERACTION WITH NEUREXOPHILIN.
RX   MEDLINE=96285495; PubMed=8699246;
RA   Petrenko A.G., Ullrich B., Missler M., Krasnoperov V., Rosahl T.W.,
RA   Suedhof T.C.;
RT   "Structure and evolution of neurexophilin.";
RL   J. Neurosci. 16:4360-4369(1996).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: The cytoplasmic C-terminal region binds to CASK. Binds
CC       SYTL1. Laminin G-like domain 2 binds to NXPH1. Isoforms 9 and
CC       alpha-4c bind to alpha-dystroglycan. Isoform alpha-4c binds to
CC       alpha-latroxin (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms, alpha-type (shown here) and
CC         beta-type, are produced by use of alternative promoters.
CC         Beta-type isoforms differ from alpha-type isoforms in their
CC         N-terminus;
CC       Event=Alternative splicing; Named isoforms=12;
CC         Comment=Additional isoforms seem to exist. There is probably
CC         more than 96 isoforms. There is a combination of five
CC         alternatively spliced domains at sites 1 to 5, each consisting
CC         of modular sequences (A-G) that seem to be used independently.
CC         For splice site 1 additional splice modules might be possible
CC         thus increasing the number of possible isoforms. Beta isoforms
CC         (AC Q28142) share the combination of two alternatively spliced
CC         domains at sites 4 and 5. Experimental confirmation may be
CC         lacking for some isoforms;
CC       Name=1; Synonyms=Alpha-1A2A3A4A5A;
CC         IsoId=Q28146-1; Sequence=Displayed;
CC       Name=2; Synonyms=Alpha-1B;
CC         IsoId=Q28146-2; Sequence=VSP_003478;
CC       Name=3; Synonyms=Alpha-1C;
CC         IsoId=Q28146-3; Sequence=VSP_003474, VSP_003478;
CC       Name=4; Synonyms=Alpha-1D;
CC         IsoId=Q28146-4; Sequence=VSP_003477;
CC       Name=5; Synonyms=Alpha-1E;
CC         IsoId=Q28146-5; Sequence=VSP_003474, VSP_003477;
CC       Name=6; Synonyms=Alpha-1F;
CC         IsoId=Q28146-6; Sequence=VSP_003476, VSP_003478;
CC       Name=7; Synonyms=Alpha-1G;
CC         IsoId=Q28146-7; Sequence=VSP_003475;
CC       Name=8; Synonyms=Alpha-2B;
CC         IsoId=Q28146-8; Sequence=VSP_003480;
CC       Name=9; Synonyms=Alpha-2C;
CC         IsoId=Q28146-9; Sequence=VSP_003479;
CC       Name=10; Synonyms=Alpha-3B;
CC         IsoId=Q28146-10; Sequence=VSP_003481;
CC       Name=11; Synonyms=Alpha-4B;
CC         IsoId=Q28146-11; Sequence=VSP_003482;
CC       Name=12; Synonyms=Alpha-5B;
CC         IsoId=Q28146-12; Sequence=VSP_003483;
CC   -!- PTM: N- and O-glycosylated (By similarity).
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
CC
DR   EMBL; L14855; AAA74123.1; -.
DR   PIR; I45944; I45944.
DR   HSSP; Q63373; 1C4R.
```

```
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 6.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00282; LamG; 6.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage.
FT   SIGNAL        1     30       BY SIMILARITY.
FT   CHAIN        31   1530       NEUREXIN 1-ALPHA.
FT   DOMAIN       31   1454       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1455   1475       POTENTIAL.
FT   DOMAIN     1476   1530       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    212       LAMININ G-LIKE 1.
FT   DOMAIN      213    256       EGF-LIKE 1.
FT   DOMAIN      299    496       LAMININ G-LIKE 2.
FT   DOMAIN      503    695       LAMININ G-LIKE 3.
FT   DOMAIN      699    736       EGF-LIKE 2.
FT   DOMAIN      741    914       LAMININ G-LIKE 4.
FT   DOMAIN      928   1103       LAMININ G-LIKE 5.
FT   DOMAIN     1106   1143       EGF-LIKE 3.
FT   DOMAIN     1149   1347       LAMININ G-LIKE 6.
FT   DOMAIN     1377   1380       POLY-THR.
FT   DOMAIN     1462   1465       POLY-ALA.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    813    813       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1246   1246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    258    258       Missing (in isoform 3 and isoform 5).
FT                                /FTId=VSP_003474.
FT   VARSPLIC    258    293       Missing (in isoform 7).
FT                                /FTId=VSP_003475.
FT   VARSPLIC    264    273       Missing (in isoform 6).
FT                                /FTId=VSP_003476.
FT   VARSPLIC    274    293       Missing (in isoform 4 and isoform 5).
FT                                /FTId=VSP_003477.
FT   VARSPLIC    278    293       Missing (in isoform 2, isoform 3 and
FT                                isoform 6).
FT                                /FTId=VSP_003478.
FT   VARSPLIC    395    409       Missing (in isoform 9).
FT                                /FTId=VSP_003479.
FT   VARSPLIC    403    409       Missing (in isoform 8).
FT                                /FTId=VSP_003480.
FT   VARSPLIC    806    815       DCIRINCNSS -> G (in isoform 10).
FT                                /FTId=VSP_003481.
FT   VARSPLIC   1263   1292       Missing (in isoform 11).
FT                                /FTId=VSP_003482.
FT   VARSPLIC   1426   1428       Missing (in isoform 12).
FT                                /FTId=VSP_003483.
SQ   SEQUENCE   1530 AA;  167937 MW;  8A4E4A75C4EC03D1 CRC64;

  Query Match             10.6%;  Score 405.5;  DB 1;  Length 1530;
  Best Local Similarity   20.5%;  Pred. No. 4.9e-18;
  Matches  203;  Conservative  112;  Mismatches  285;  Indels  391;  Gaps   42;

Qy          2 IKITFRPDSADGMLLY-NGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              |   ||    :|::|: :|: |      :    : || :  :: |      | |||   |
Db        527 ISFDFRTTEPNGLILFSHGKPRHQKDAKHPQMIKVDFFAIEMLDGHLYLLLDMGSGTIKI 586

Qy         61 RHPTPLAL------GHFHTVTLLRSLTQGSLIVGDL-APVNGTSQGKFQGLDLNEELYLG 113
              :     ||      | :: |   |    |:: |  |  |   |: |: | |||::|||||
Db        587 K-----ALQKKVNDGEWYHVDFQRDGRSGTISVNTLRTPY--TAPGESQILDLDDELYLG 639
```

```
Qy        114 GYPDYGAIPKAG-----------LSSGFIGCVRELRIQGEEIVFH---DLNLTAHGISHC 159
              | |:     |||           |: |::||:|:| | |:        ::  ||     |
Db        640 GLPE----NKAGLVFPTEVWTALLNYGYVGCIRDLFIDGQSKDIRQMAEVQSTAGVKPSC 695

Qy        160 -----PTCRDRPCQNGGQCHDSESSSYVCVCP-AGFTGSRCE------------------ 195
                     |   ||:| | | |   : ||| |   |: |  ||
Db        696 SRETAKPCLSNPCKNNGMCRDG-WNRYVCDCSGTGYLGRSCEREATVLSYDGSMFMKIQL 754

Qy        196 ----HSQA------------------------------------------LHCHPEAC-- 207
                  |::|                                          | |    |
Db        755 PVVMHTEAEDVSLRFRSQRAYGILMATTSRDSADTLRLELDAGRVKLTVNLDCIRINCNS 814

Qy        208 --GPDATCV--NRPDGRGYTCR-------------------------------------- 225
                ||:      |  |   :| |
Db        815 SKGPETLFAGYNLNDNEWHTVRVVRRGKSLKLTVDDQQAMTGQMAGDHTRLEFHNIETGI 874

Qy        226 ---------------------------------------CHL-GRSGLR--CEEGVTVTT 243
                                                     | |  | | |    : ||  |
Db        875 ITERRYLSSVPSNFIGHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIIADPVTFKT 934

Qy        244 PSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHL 303
               |     ||:||  |   :  : |  :||  : ||::|::   ||   ||: : :| |:|
Db        935 KS-----SYVALATL-QAYTSMHLFFQFKTTSLDGLILYN---SGDGNDFIVVELVKGYL 985

Qy        304 EFRYELGSGLAVLR--SAEPLALGRWHRVSAER---------LNKDGSLRVNGGRPVLRS 352
               : ::||:|  :::  | :||   :|| |   |         ::   : ::  |
Db        986 HYVFDLGNGANLIKGSSNKPLNDNQWHNVMISRDTSNLHTVKIDTKITTQITAG------ 1039

Qy        353 SPGKSQGLNLHTLLYLGGV--EPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGS 410
                  :: |:| : ||:|||  |    |    :    |:||:  | :||:  ||
Db       1040 ----ARNLDLKSDLYIGGVAKETYKSLPKLVHAKEGFQGCLASVDLNGRLPDL------- 1088

Qy        411 QGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHG 470
                     |:  |              |: |         :|  ||     ||  : | :
Db       1089 -----ISDALFCN-------------GQIE---------RG--CEGPSTTCQ-EDSCSNQ 1118

Qy        471 GTC----QGTRCLC-LPGFSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGA 518
              | |     |  | | :  |||| |           | |     |       || |
Db       1119 GVCLQQWDGISCDCSMTSFSGPLCNDPGTTYIFSKGGGQITYKW-----PPNDRPSTRA- 1172

Qy        519 YFHDDGFLAFPGHVFSRSLPEVPETIELE---VRTSTASGLLLWQGVEVGEAGQGKDFIS 575
                 |   : |              |:: |   ||  ::|||               |::
Db       1173 ---DRLAIGF-------------STVQKEAVLVRVDSSSGL--------------GDYLE 1202

Qy        576 LGLQDGHLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDG----EELVS 630
              | :  | :  :: :|: :  :   :  ||||::| |   | |   ::|||     |   :
Db       1203 LHIHQGKIGVKFNVGTDDIAIEESNAIINDGKYHVVRFTRSGGNATLQVDSWPVIERYPA 1262

Qy        631 GRSPGPNVAV---------------------------NAKGSVYIGG------------- 650
              | :    :|:                           |::  :: |||
Db       1263 GNNDNERLAIARQRIPYRLGRVVDEWLLDKGRQLTIFNSQATIIIGGKEQGQPFQGQLSG 1322

Qy        651 ------------------------------------APDVATLTGGRFSSGITGCVKNLV 674
                                                      ||      |: |      |
Db       1323 LYYNGLKVLNMAAENDANIAIVGNVRLVGEVPSSMTTESTATAMQSEMSTSIMETTTTLA 1382

Qy        675 LHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                :|| | || :    |        :  |||
Db       1383 TSTARRGKPPTKEPVSQTTDDILVASAECPS 1413


RESULT 12
NX1A_RAT
ID   NX1A_RAT        STANDARD;      PRT;  1514 AA.
AC   Q63372;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 1-alpha precursor (Neurexin I-alpha).
GN   NRXN1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

```
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=92320296; PubMed=1621094;
RA   Ushkaryov Y.A., Petrenko A.G., Geppert M., Suedhof T.C.;
RT   "Neurexins: synaptic cell surface proteins related to the alpha-
RT   latrotoxin receptor and laminin.";
RL   Science 257:50-56(1992).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=95209856; PubMed=7695896;
RA   Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT   "Cartography of neurexins: more than 1000 isoforms generated by
RT   alternative splicing and expressed in distinct subsets of neurons.";
RL   Neuron 14:497-507(1995).
RN   [3]
RP   SEQUENCE OF N-TERMINUS, AND INTERACTION WITH NEUREXOPHILINS 1 AND 3.
RX   MEDLINE=99074239; PubMed=9856994;
RA   Missler M., Hammer R.E., Suedhof T.C.;
RT   "Neurexophilin binding to alpha-neurexins. A single LNS domain
RT   functions as an independently folding ligand-binding unit.";
RL   J. Biol. Chem. 273:34716-34723(1998).
RN   [4]
RP   INTERACTION WITH CASK.
RX   MEDLINE=96256685; PubMed=8786425;
RA   Hata Y., Butz S., Suedhof T.C.;
RT   "CASK: a novel dlg/PSD95 homolog with an N-terminal
RT   calmodulin-dependent protein kinase domain identified by interaction
RT   with neurexins.";
RL   J. Neurosci. 16:2488-2494(1996).
RN   [5]
RP   INTERACTION WITH ALPHA-LATROTOXIN.
RX   MEDLINE=99211386; PubMed=10197529;
RA   Sugita S., Khvochtev M., Suedhof T.C.;
RT   "Neurexins are functional alpha-latrotoxin receptors.";
RL   Neuron 22:489-496(1999).
RN   [6]
RP   INTERACTION WITH ALPHA-DYSTROGLYCAN.
RX   MEDLINE=21363578; PubMed=11470830;
RA   Sugita S., Saito F., Tang J., Satz J., Campbell K., Suedhof T.C.;
RT   "A stoichiometric complex of neurexins and dystroglycan in brain.";
RL   J. Cell Biol. 154:435-445(2001).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: Binds SYTL1 (By similarity). The cytoplasmic C-terminal
CC       region binds to CASK. Laminin G-like domain 2 binds to NXPH1 and
CC       NXPH3. Isoforms alpha 2C and alpha 4C bind to alpha-dystroglycan.
CC       Isoforms alpha 4C bind to alpha-latroxin.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative promoter;
CC         Comment=A number of isoforms, alpha-type (shown here) and
CC         beta-type, are produced by use of alternative promoters.
CC         Beta-type isoforms differ from alpha-type isoforms in their
CC         N-terminus;
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=At least 96 isoforms may be produced by alternative
CC         splicing. There is a combination of five alternatively spliced
CC         domains at sites 1 to 5, each consisting of modular sequences
CC         (A-G) that seem to be used independently. For splice site 1
CC         additional splice modules might be possible thus increasing the
CC         number of possible isoforms. Beta-type isoforms (AC Q63373)
CC         share the combination of two alternatively spliced domains at
CC         sites 4 and 5;
CC       Name=Alpha 1A2A3A4A5A;
CC         IsoId=Q63372-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Brain (neuronal synapse).
CC   -!- PTM: N- and O-glycosylated (By similarity).
CC   -!- MISCELLANEOUS: Alpha-latroxin competes with alpha-dystroglycan for
CC       binding.
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
```

```
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M96374; AAA41704.1; -.
DR   PIR; A40228; A40228.
DR   HSSP; Q63373; 1C4R.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 6.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 5.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing; Alternative promoter usage.
FT   SIGNAL        1     30
FT   CHAIN        31   1514       NEUREXIN 1-ALPHA.
FT   DOMAIN       31   1438       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1439   1459       POTENTIAL.
FT   DOMAIN     1460   1514       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    217       LAMININ G-LIKE 1.
FT   DOMAIN      213    255       EGF-LIKE 1.
FT   DOMAIN      283    480       LAMININ G-LIKE 2.
FT   DOMAIN      487    679       LAMININ G-LIKE 3.
FT   DOMAIN      683    720       EGF-LIKE 2.
FT   DOMAIN      725    898       LAMININ G-LIKE 4.
FT   DOMAIN      912   1087       LAMININ G-LIKE 5.
FT   DOMAIN     1090   1127       EGF-LIKE 3.
FT   DOMAIN     1133   1331       LAMININ G-LIKE 6.
FT   DOMAIN     1361   1364       POLY-THR.
FT   DOMAIN     1446   1449       POLY-ALA.
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    797    797       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1230   1230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    258    258       Missing (in some isoforms Alpha 1C).
FT                                /FTId=VSP_003486.
FT   VARSPLIC    258    277       Missing (in some isoforms Alpha 1G).
FT                                /FTId=VSP_003487.
FT   VARSPLIC    274    277       Missing (in some isoforms Alpha 1D).
FT                                /FTId=VSP_003488.
FT   VARSPLIC    379    393       Missing (in some isoforms Alpha 2C).
FT                                /FTId=VSP_003489.
FT   VARSPLIC    387    393       Missing (in some isoforms Alpha 2B).
FT                                /FTId=VSP_003490.
FT   VARSPLIC    790    799       DCIRINCNSS -> G (in some isoforms Alpha
FT                                3B).
FT                                /FTId=VSP_003491.
FT   VARSPLIC   1247   1276       Missing (in some isoforms Alpha 4B).
FT                                /FTId=VSP_003492.
FT   VARSPLIC   1410   1411       Missing (in some isoforms Alpha 5B).
FT                                /FTId=VSP_003493.
SQ   SEQUENCE   1514 AA;  166167 MW;  D99492302968DE14 CRC64;

  Query Match             10.5%;  Score 403.5;  DB 1;  Length 1514;
  Best Local Similarity   20.4%;  Pred. No. 6.5e-18;
```

```
  Matches  202;  Conservative  113;  Mismatches  285;  Indels  391;  Gaps   42;

Qy           2 IKITFRPDSADGMLLY-NGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
               |   ||    :|::|: :|: |      :    : || :  :: |      | |||   |
Db         511 ISFDFRTTEPNGLILFSHGKPRHQKDAKHPQMIKVDFFAIEMLDGHLYLLLDMGSGTIKI 570

Qy          61 RHPTPLAL------GHFHTVTLLRSLTQGSLIVGDL-APVNGTSQGKFQGLDLNEELYLG 113
               :     ||      | :: |   |    |:: |  |  |   |: |: : |||::|||||
Db         571 K-----ALQKKVNDGEWYHVDFQRDGRSGTISVNTLRTPY--TAPGESEILDLDDELYLG 623

Qy         114 GYPDYGAIPKAG-----------LSSGFIGCVRELRIQGEEIVFH---DLNLTAHGISHC 159
               | |:     |||           |: |::||:|:| | |:        ::  ||     |
Db         624 GLPE----NKAGLVFPTEVWTALLNYGYVGCIRDLFIDGQSKDIRQMAEIQSTAGVKPSC 679

Qy         160 -----PTCRDRPCQNGGQCHDSESSSYVCVCP-AGFTGSRCE------------------ 195
                      |   ||:| | | |   : ||| |   |: |  ||
Db         680 SRETAKPCLSNPCKNNGMCRDG-WNRYVCDCSGTGYLGRSCEREATVLSYDGSMFMKIQL 738

Qy         196 ----HSQA------------------------------------------LHCHPEAC-- 207
                   |::|                                          | |    |
Db         739 PVVMHTEAEDVSLRFRSQRAYGILMATTSRDSADTLRLELDAGRVKLTVNLDCIRINCNS 798

Qy         208 --GPDATCV--NRPDGRGYTCR-------------------------------------- 225
                 ||:      |  |   :| |
Db         799 SKGPETLFAGYNLNDNEWHTVRVVRRGKSLKLTVDDQQAMTGQMAGDHTRLEFHNIETGI 858

Qy         226 ---------------------------------------CHL-GRSGLR--CEEGVTVTT 243
                                                      | |  | | |    : ||  |
Db         859 ITERRYLSSVPSNFIGHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIIADPVTFKT 918

Qy         244 PSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHL 303
                |     ||:||  |   :  : |  :||  : ||::|::   ||   ||: : :| |:|
Db         919 KS-----SYVALATL-QAYTSMHLFFQFKTTSLDGLILYN---SGDGNDFIVVELVKGYL 969

Qy         304 EFRYELGSGLAVLR--SAEPLALGRWHRVSAER---------LNKDGSLRVNGGRPVLRS 352
                : ::||:|  :::  | :||   :|| |   |         ::   : ::  |
Db         970 HYVFDLGNGANLIKGSSNKPLNDNQWHNVMISRDTSNLHTVKIDTKITTQITAG------ 1023

Qy         353 SPGKSQGLNLHTLLYLGGV--EPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGS 410
                   :: |:| : ||:|||  |    |    :    |:||:  | :||:  ||
Db        1024 ----ARNLDLKSDLYIGGVAKETYKSLPKLVHAKEGFQGCLASVDLNGRLPDL------- 1072

Qy         411 QGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHG 470
                      |:  |              |: |         :|  ||     ||  : | :
Db        1073 -----ISDALFCN-------------GQIE---------RG--CEGPSTTCQ-EDSCSNQ 1102

Qy         471 GTC----QGTRCLC-LPGFSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGA 518
               | |     |  | | :  |||| |           | |     |       || |
Db        1103 GVCLQQWDGFSCDCSMTSFSGPLCNDPGTTYIFSKGGGQITHKW-----PPNDRPSTRA- 1156

Qy         519 YFHDDGFLAFPGHVFSRSLPEVPETIELE---VRTSTASGLLLWQGVEVGEAGQGKDFIS 575
                  |   : |              |:: |   ||  ::|||               |::
Db        1157 ---DRLAIGF-------------STVQKEAVLVRVDSSSGL--------------GDYLE 1186

Qy         576 LGLQDGHLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDG----EELVS 630
               | :  | :  :: :|: :  :   :  ||||::| |   | |   ::|||     |   :
Db        1187 LHIHQGKIGVKFNVGTDDIAIEESNAIINDGKYHVVRFTRSGGNATLQVDSWPVIERYPA 1246

Qy         631 GRSPGPNVAV---------------------------NAKGSVYIGG------------- 650
               | :    :|:                           |:: :: |||
Db        1247 GNNDNERLAIARQRIPYRLGRVVDEWLLDKGRQLTIFNSQATIIIGGKEQGQPFQGQLSG 1306

Qy         651 ------------------------------------APDVATLTGGRFSSGITGCVKNLV 674
                                                       ||      |: |      |
Db        1307 LYYNGLKVLNMAAENDANIAIVGNVRLVGEVPSSMTTESTATAMQSEMSTSIMETTTTLA 1366

Qy         675 LHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                 :|| | || :    |        :  |||
Db        1367 TSTARRGKPPTKEPISQTTDDILVASAECPS 1397


RESULT 13
CLR3_MOUSE
```

```
ID   CLR3_MOUSE      STANDARD;      PRT;  3301 AA.
AC   Q91ZI0; Q9ESD0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cadherin EGF LAG seven-pass G-type receptor 3 precursor.
GN   CELSR3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGES.
RC   STRAIN=C57BL/6;
RX   MEDLINE=21839555; PubMed=11850187;
RA   Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT   "Developmental expression profiles of Celsr (Flamingo) genes in the
RT   mouse.";
RL   Mech. Dev. 112:157-160(2002).
RN   [2]
RP   SEQUENCE OF 2574-3046 FROM N.A., AND DEVELOPMENTAL STAGE.
RX   MEDLINE=21534880; PubMed=11677057;
RA   Formstone C.J., Little P.F.R.;
RT   "The flamingo-related mouse Celsr family (Celsr1-3) genes exhibit
RT   distinct patterns of expression during embryonic development.";
RL   Mech. Dev. 109:91-94(2001).
RN   [3]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=20253755; PubMed=10790539;
RA   Formstone C.J., Barclay J., Rees M., Little P.F.R.;
RT   "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is
RT   a candidate for the tippy (tip) lethal mutant on chromosome 9.";
RL   Mamm. Genome 11:392-394(2000).
CC   -!- FUNCTION: Receptor that may have an important role in cell/cell
CC       signaling during nervous system formation.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.
CC   -!- DEVELOPMENTAL STAGE: Predominantly expressed in the CNS, the
CC       emerging dorsal root ganglia and cranial ganglia. In the CNS,
CC       expression is uniform along the rostrocaudal axis. No expression
CC       is detected until somite stages. Between E10 and E12, expression
CC       is strong in the marginal zone (MZ), and lower in the ventricular
CC       zone (VZ). At E15, expression is restricted to the brain and
CC       olfactory epithelium. In the brain, it is low in VZ but strong in
CC       external fields, particularly those with ongoing migration, such
CC       as the telencephalic cortical plate, the olfactory bulb, the
CC       cerebrellum and the tectum. In the newborn and postnatal stages,
CC       expression is high in differenciated neuronal fields.
CC   -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC   -!- SIMILARITY: Contains 9 cadherin domains.
CC   -!- SIMILARITY: Contains 7 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 laminin EGF-like domain.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC       in positions 2575 and 2578.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF427498; AAL25099.1; -.
DR   EMBL; AF188752; AAG17057.1; ALT_FRAME.
DR   MGD; MGI:1858236; Celsr3.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 9.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00053; laminin_EGF; 1.
DR   Pfam; PF00054; laminin_G; 2.
DR   PRINTS; PR00205; CADHERIN.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00112; CA; 9.
DR   SMART; SM00181; EGF; 6.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00232; CADHERIN_1; 7.
DR   PROSITE; PS50268; CADHERIN_2; 8.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW   Developmental protein; Hydroxylation; Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32   3301       CADHERIN EGF LAG SEVEN-PASS G-TYPE
FT                                RECEPTOR 3.
FT   DOMAIN       32   2531       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2532   2552       1 (POTENTIAL).
FT   DOMAIN     2553   2563       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2564   2584       2 (POTENTIAL).
FT   DOMAIN     2585   2592       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2593   2613       3 (POTENTIAL).
FT   DOMAIN     2614   2634       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2635   2655       4 (POTENTIAL).
FT   DOMAIN     2656   2673       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2674   2694       5 (POTENTIAL).
FT   DOMAIN     2695   2716       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2717   2737       6 (POTENTIAL).
FT   DOMAIN     2738   2744       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2745   2765       7 (POTENTIAL).
FT   DOMAIN     2766   3301       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      317    424       CADHERIN 1.
FT   DOMAIN      425    536       CADHERIN 2.
FT   DOMAIN      537    642       CADHERIN 3.
FT   DOMAIN      643    747       CADHERIN 4.
FT   DOMAIN      748    849       CADHERIN 5.
FT   DOMAIN      850    952       CADHERIN 6.
FT   DOMAIN      953   1058       CADHERIN 7.
FT   DOMAIN     1059   1160       CADHERIN 8.
FT   DOMAIN     1161   1257       CADHERIN 9.
FT   DOMAIN     1366   1424       EGF-LIKE 1, CALCIUM-BINDING.
FT   DOMAIN     1426   1460       EGF-LIKE 2, CALCIUM-BINDING.
FT   DOMAIN     1464   1503       EGF-LIKE 3, CALCIUM-BINDING.
FT   DOMAIN     1504   1708       LAMININ G-LIKE 1.
FT   DOMAIN     1711   1747       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN     1751   1933       LAMININ G-LIKE 2.
FT   DOMAIN     1935   1971       EGF-LIKE 5, CALCIUM-BINDING.
FT   DOMAIN     1972   2002       EGF-LIKE 6, CALCIUM-BINDING.
FT   DOMAIN     2003   2042       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN     2044   2079       EGF-LIKE 8, CALCIUM-BINDING.
```

```
FT   DOMAIN     2085   2118       LAMININ EGF-LIKE.
FT   DOMAIN     2468   2520       GPS.
FT   DOMAIN     2720   2724       POLY-LEU.
FT   DISULFID   1370   1381       BY SIMILARITY.
FT   DISULFID   1375   1412       BY SIMILARITY.
FT   DISULFID   1414   1423       BY SIMILARITY.
FT   DISULFID   1430   1441       BY SIMILARITY.
FT   DISULFID   1435   1450       BY SIMILARITY.
FT   DISULFID   1452   1459       BY SIMILARITY.
FT   DISULFID   1468   1479       BY SIMILARITY.
FT   DISULFID   1473   1489       BY SIMILARITY.
FT   DISULFID   1491   1502       BY SIMILARITY.
FT   DISULFID   1715   1726       BY SIMILARITY.
FT   DISULFID   1720   1735       BY SIMILARITY.
FT   DISULFID   1737   1746       BY SIMILARITY.
FT   DISULFID   1939   1950       BY SIMILARITY.
FT   DISULFID   1944   1959       BY SIMILARITY.
FT   DISULFID   1961   1970       BY SIMILARITY.
FT   DISULFID   1974   1985       BY SIMILARITY.
FT   DISULFID   1979   1997       BY SIMILARITY.
FT   DISULFID   1999   2008       BY SIMILARITY.
FT   DISULFID   2016   2029       BY SIMILARITY.
FT   DISULFID   2031   2041       BY SIMILARITY.
FT   DISULFID   2048   2063       BY SIMILARITY.
FT   DISULFID   2050   2066       BY SIMILARITY.
FT   DISULFID   2068   2078       BY SIMILARITY.
FT   MOD_RES    1952   1952       HYDROXYLATION (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    838    838       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1173   1173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1213   1213       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1308   1308       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1318   1318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1638   1638       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1702   1702       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1759   1759       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2042   2042       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2166   2166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2185   2185       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2375   2375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2465   2465       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2497   2497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   2713   2713       L -> LR (IN REF. 2).
FT   CONFLICT   3024   3024       R -> P (IN REF. 2).
SQ   SEQUENCE   3301 AA;  358455 MW;  A6B18F2DF7F4DEB6 CRC64;

  Query Match             10.5%;  Score 403;  DB 1;  Length 3301;
  Best Local Similarity   23.0%;  Pred. No. 1.7e-17;
  Matches  179;  Conservative   91;  Mismatches  263;  Indels  244;  Gaps   40;

Qy          61 RHPTPLALGHF--HTVTLLRSLTQGSLI--------VGD---------LAPVNGTSQGK- 100
               |  :|| ||||      :| : |:   |        ||          |||  |   |
Db        1277 RFLSPL-LGHFLEGVAAVLATPTEDVFIFNIQNDTDVGGTVLNVSFSALAP-RGAGAGAA 1334

Qy         101 ---FQGLDLNEELYLGGYPDYGAIPKAGLSS--------------------GFIGCVREL 137
                  |   :| |:||         : :| |::                     :: ||  |
Db        1335 GPWFSSEELQEQLY---------VRRAALAARSLLDVLPFDDNVCLREPCENYMKCVSVL 1385

Qy         138 R-------IQGEEIVFHDLNLTAHGISHCP-------------TCRDRPCQNGGQCHDSE 177
               |       :     :|  :   |     ||              |   ||:||| |   |
Db        1386 RFDSSAPFLASTSTLFRPIQPIAGLRCRCPPGFTGDFCETELDLCYSNPCRNGGACARRE 1445

Qy         178 SSSYVCVCPAGFTGSRCE-HSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRS--GLR 234
                  | |||   ||   ||  ::|  | |  |    || | |:| |: |:|  | :  | |
Db        1446 -GGYTCVCRPRFTD--CELDTEAGRCVPGVCRNGGTCTNAPNG-GFRCQCPAGGAFEGPR 1501

Qy         235 CEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFV 294
               ||       ||     |::    |    | | | : |  : | | |||  |:     ||:
Db        1502 CEVAARSFPPS-----SFVMFRGLRQRFH-LTLSLSFATVQPSG-LLFYNGRLNEKHDFL 1554

Qy         295 SLAMVGGHLEFRYELGSGLAVLRSAEP--LALGRWHRVSAERLNKDGSLRVNGGRPVLRS 352
               :| :| | :   |  |    |:    |  |: |:|| |     ||  :  : | :
Db        1555 ALELVAGQVRLTYSTGESNTVVSPTVPGGLSDGQWHTVHLRYYNKPRTDALGGAQ----- 1609

Qy         353 SPGK-------------------------------------SQGLNLHTLLYLGGVEPSV 375
                | |                                      : |:|   | |||| |::
Db        1610 GPSKDKVAVLSVDDCNVAVALQFGAEIGNYSCAAAGVQTSSKKSLDLTGPLLLGGV-PNL 1668

Qy         376 PLS-PATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCM 434
               | : | ::    | ||: :: ::|:|:|:  :|: : |                  | |
Db        1669 PENFPVSHKD--FIGCMRDLHIDGRRMDMA-AFVANNGT----------------MAGCQ 1709

Qy         435 PAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ----GTRCLCLPGFSGPRCQ 490
                   :     |  |                 || : | |     |  | |  || |  |:
Db        1710 AKSHF-----CASG-----------------PCKNNGFCSERWGGFSCDCPVGFGGKDCR 1747

Qy         491 QGSGHGIAESDWHLEGSG------GNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETI 544
                   |      :| :|:|      |||                            ||  :
Db        1748 LTMAH-----PYHFQGNGTLSWDFGND------------------------MAVSVPWYL 1778

Qy         545 ELEVRTSTASGLLLWQGVEVG-------EAGQGKDFISLGLQDGHLVFRYQLGSGEARLV 597
                |  ||    |:|:   |::|       :  :|   ::|    || |           |:
Db        1779 GLSFRTRATKGILM--QVQLGPHSVLLCKLDRGLLSVTLNRASGHTVH---------LLL 1827

Qy         598 SEDPINDGEWH--RVTALRE--GRRG----SIQVDGEELVSGRSPGPNVAVNAKGSVYIG 649
                :  ::|| ||  |:    |  ||||     : :|        : |  :       :::|
Db        1828 DQMTVSDGRWHDLRLELQEEPGGRRGHHIFMVSLDFTLFQDTMAMGGELQGLKVKQLHVG 1887

Qy         650 GAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGAN-TRPCPS 705
               | |  :   |     |: ||:: : :     |:    |   :   : |   | ||  |
Db        1888 GLPPSSKEEG---HQGLVGCIQGVWIGFTPFGSSALLPPSHRVNVEPGCTVTNPCAS 1941


RESULT 14
SLT2_HUMAN
ID   SLT2_HUMAN     STANDARD;      PRT;  1529 AA.
AC   O94813; O95710; Q9Y5Q7;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Slit homolog 2 protein precursor (h-Slit-2).
GN   SLIT2 OR SLIL2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC   TISSUE=Fetal lung;
RX   MEDLINE=99033071; PubMed=9813312;
RA   Itoh A., Miyabayashi T., Ohno M., Sakano S.;
RT   "Cloning and expressions of three mammalian homologues of Drosophila
RT   slit suggest possible roles for slit in the formation and maintenance
RT   of the nervous system.";
RL   Brain Res. Mol. Brain Res. 62:175-186(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 3), AND TISSUE SPECIFICITY.
RC   TISSUE=Fetal brain, and Fetal kidney;
RX   MEDLINE=99279238; PubMed=10349621;
RA   Holmes G.P., Negus K., Burridge L., Raman S., Algar E., Yamada T.,
RA   Little M.H.;
RT   "Distinct but overlapping expression patterns of two vertebrate slit
RT   homologs implies functional roles in CNS development and
RT   organogenesis.";
RL   Mech. Dev. 79:57-72(1998).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 1122-1129, FUNCTION,
RP   SUBUNIT, AND SUBCELLULAR LOCATION.
RC   TISSUE=Fetal brain;
RX   MEDLINE=99200391; PubMed=10102268;
RA   Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA   Tessier-Lavigne M., Kidd T.;
RT   "Slit proteins bind Robo receptors and have an evolutionarily
RT   conserved role in repulsive axon guidance.";
RL   Cell 96:795-806(1999).
CC   -!- FUNCTION: Plays a role in axon guidance as a repulsive ligand for
```

```
CC       Robo receptors preventing inappropriate midline crossing.
CC   -!- SUBUNIT: Binds robo proteins with high affinity.
CC   -!- SUBCELLULAR LOCATION: Secreted. The C-terminal cleavage protein is
CC       more diffusible than the larger N-terminal protein that is more
CC       tightly cell associated.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=O94813-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O94813-2; Sequence=VSP_050035, VSP_050036;
CC       Name=3;
CC         IsoId=O94813-3; Sequence=VSP_050036;
CC   -!- TISSUE SPECIFICITY: Fetal lung and kidney, and adult spinal cord.
CC       Weak expression in adult adrenal gland, thyroid, trachea and other
CC       tissues examined.
CC   -!- SIMILARITY: Contains 7 EGF-like domains.
CC   -!- SIMILARITY: Contains 23 leucine-rich (LRR) repeats.
CC   -!- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.
CC   -!- SIMILARITY: Contains 1 laminin G-like domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB017168; BAA35185.1; -.
DR   EMBL; AF055585; AAD04309.1; -.
DR   EMBL; AF133270; AAD25539.1; -.
DR   HSSP; P00743; 1CCF.
DR   Genew; HGNC:11086; SLIT2.
DR   MIM; 603746; -.
DR   GO; GO:0005576; C:extracellular; IEP.
DR   GO; GO:0045499; F:chemorepellant activity; IMP.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0008347; P:glia cell migration; IMP.
DR   GO; GO:0007509; P:mesoderm migration; IMP.
DR   GO; GO:0008045; P:motor axon guidance; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   GO; GO:0008038; P:neuronal cell recognition; TAS.
DR   GO; GO:0007608; P:olfaction; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00008; EGF; 8.
DR   Pfam; PF00054; laminin_G; 1.
DR   Pfam; PF00560; LRR; 18.
DR   Pfam; PF01463; LRRCT; 4.
DR   Pfam; PF01462; LRRNT; 4.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00274; FOLN; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00369; LRR_TYP; 8.
DR   SMART; SM00082; LRRCT; 4.
DR   SMART; SM00013; LRRNT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS50026; EGF_3; 9.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Neurogenesis; Glycoprotein; Signal; Alternative splicing;
KW   EGF-like domain; Repeat; Leucine-rich repeat.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1529       SLIT PROTEIN.
FT   CHAIN        31   1121       SLIT PROTEIN N-PRODUCT.
FT   CHAIN      1122   1529       SLIT PROTEIN C-PRODUCT.
FT   SITE       1121   1122       CLEAVAGE.
FT   REPEAT       54     77       LRR 1.
FT   REPEAT       78    101       LRR 2.
FT   REPEAT      102    125       LRR 3.
FT   REPEAT      126    149       LRR 4.
FT   REPEAT      151    173       LRR 5.
FT   REPEAT      174    197       LRR 6.
FT   REPEAT      199    226       LRR 7.
FT   REPEAT      299    322       LRR 8.
FT   REPEAT      323    346       LRR 9.
FT   REPEAT      347    370       LRR 10.
FT   REPEAT      372    394       LRR 11.
FT   REPEAT      395    418       LRR 12.
FT   REPEAT      511    531       LRR 13.
FT   REPEAT      532    556       LRR 14.
FT   REPEAT      557    580       LRR 15.
FT   REPEAT      582    604       LRR 16.
FT   REPEAT      605    628       LRR 17.
FT   REPEAT      629    652       LRR 18.
FT   REPEAT      753    776       LRR 19.
FT   REPEAT      777    799       LRR 20.
FT   REPEAT      800    823       LRR 21.
FT   REPEAT      824    847       LRR 22.
FT   REPEAT      849    872       LRR 23.
FT   DOMAIN      918    955       EGF-LIKE 1.
FT   DOMAIN      957    996       EGF-LIKE 2.
FT   DOMAIN      998   1034       EGF-LIKE 3, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN     1036   1074       EGF-LIKE 4.
FT   DOMAIN     1076   1112       EGF-LIKE 5, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN     1121   1157       EGF-LIKE 6.
FT   DOMAIN     1160   1333       LAMININ G-LIKE.
FT   DOMAIN     1332   1368       EGF-LIKE 7.
FT   DOMAIN     1453   1528       CTCK.
FT   DISULFID    922    933       BY SIMILARITY.
FT   DISULFID    927    943       BY SIMILARITY.
FT   DISULFID    945    954       BY SIMILARITY.
FT   DISULFID    961    972       BY SIMILARITY.
FT   DISULFID    966    984       BY SIMILARITY.
FT   DISULFID    986    995       BY SIMILARITY.
FT   DISULFID   1002   1013       BY SIMILARITY.
FT   DISULFID   1007   1022       BY SIMILARITY.
FT   DISULFID   1024   1033       BY SIMILARITY.
FT   DISULFID   1040   1053       BY SIMILARITY.
FT   DISULFID   1047   1062       BY SIMILARITY.
FT   DISULFID   1064   1073       BY SIMILARITY.
FT   DISULFID   1080   1091       BY SIMILARITY.
FT   DISULFID   1085   1100       BY SIMILARITY.
FT   DISULFID   1102   1111       BY SIMILARITY.
FT   DISULFID   1125   1136       BY SIMILARITY.
FT   DISULFID   1130   1145       BY SIMILARITY.
FT   DISULFID   1147   1156       BY SIMILARITY.
FT   DISULFID   1336   1346       BY SIMILARITY.
FT   DISULFID   1341   1356       BY SIMILARITY.
FT   DISULFID   1358   1367       BY SIMILARITY.
FT   CARBOHYD     66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    186    186       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    794    794       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    799    799       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1009   1009       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1010   1010       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1019   1019       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1183   1183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1266   1266       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1300   1300       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    258    258       S -> SDEEE (in isoform 2).
FT                                /FTId=VSP_050035.
FT   VARSPLIC    480    487       Missing (in isoform 2 and isoform
FT                                3).
FT                                /FTId=VSP_050036.
FT   CONFLICT    226    226       Q -> K (IN REF. 2).
FT   CONFLICT    607    610       SLKT -> KPQN (IN REF. 3).
FT   CONFLICT    634    634       L -> M (IN REF. 3).
SQ   SEQUENCE   1529 AA;  169869 MW;  5D19CC5E7FD461BA CRC64;

  Query Match             10.4%;  Score 397.5;  DB 1;  Length 1529;
  Best Local Similarity   29.7%;  Pred. No. 1.6e-17;
  Matches  108;  Conservative   58;  Mismatches  141;  Indels   57;  Gaps   18;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |:|| |::|  || |  :       |    |
Db       1072 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDC 1130

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   :| |      |:|  |  | :||: |:|   :     ||| :|:      |
Db       1131 QNGAQCIVRINEP-----ICQCLPGYQGEKCEKLVSV---NFINKESYLQIPSAKVRPQT 1182

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | ::: :  | :   |: ||  |  : |
Db       1183 NITLQIATDED------SGILLYKGDK-----DHIAVELYRGRVRASYDTGSHPASAIYS 1231

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |::  || |:|| | : ::  |   ||  : ||:||:
Db       1232 VETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVA 1291

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : :  :  |    || :: | || ||
Db       1292 SLRQAPGQN-GTSFHGCIRNLYINSELQDFQKVPMQTGILPGC---EPCHKKVCAHG-TC 1346

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | |::|: | ||:   | || |   |:| |||         | || |  |
Db       1347 QPSSQAGFTCECQEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 1404

Qy        488 RCQQ 491
               | :
Db       1405 LCDE 1408

RESULT 15
CLR3_HUMAN
ID   CLR3_HUMAN     STANDARD;      PRT;  3312 AA.
AC   Q9NYQ7; O75092;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Flamingo
DE   homolog 1) (hFmi1) (Multiple epidermal growth factor-like domains 2)
DE   (Epidermal growth factor-like 1).
GN   CELSR3 OR CDHF11 OR FMI1 OR EGFL1 OR MEGF2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20202599; PubMed=10716726;
RA   Wu Q., Maniatis T.;
RT   "Large exons encoding multiple ectodomains are a characteristic
RT   feature of protocadherin genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN   [2]
RP   SEQUENCE OF 1954-3312 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98360089; PubMed=9693030;
RA   Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
CC   -!- FUNCTION: Receptor that may have an important role in cell/cell
CC       signaling during nervous system formation.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC   -!- SIMILARITY: Contains 9 cadherin domains.
CC   -!- SIMILARITY: Contains 8 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 laminin EGF-like domain.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF231023; AAF61929.1; -.
DR   EMBL; AB011536; BAA32464.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:3230; CELSR3.
DR   MIM; 604264; -.
DR   GO; GO:0005198; F:structural molecule activity; NAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 9.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00054; laminin_G; 2.
DR   PRINTS; PR00205; CADHERIN.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00112; CA; 9.
DR   SMART; SM00181; EGF; 6.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00232; CADHERIN_1; 7.
DR   PROSITE; PS50268; CADHERIN_2; 8.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW   Developmental protein; Hydroxylation; Signal.
FT   SIGNAL         1     32       POTENTIAL.
FT   CHAIN         33   3312       CADHERIN EGF LAG SEVEN-PASS G-TYPE
FT                                 RECEPTOR 3.
```

```
FT   DOMAIN        33   2540       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    2541   2561       1 (POTENTIAL).
FT   DOMAIN      2562   2572       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    2573   2593       2 (POTENTIAL).
FT   DOMAIN      2594   2601       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    2602   2622       3 (POTENTIAL).
FT   DOMAIN      2623   2643       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    2644   2664       4 (POTENTIAL).
FT   DOMAIN      2665   2681       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    2682   2702       5 (POTENTIAL).
FT   DOMAIN      2703   2725       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    2726   2746       6 (POTENTIAL).
FT   DOMAIN      2747   2753       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    2754   2774       7 (POTENTIAL).
FT   DOMAIN      2775   3312       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       326    433       CADHERIN 1.
FT   DOMAIN       434    545       CADHERIN 2.
FT   DOMAIN       546    651       CADHERIN 3.
FT   DOMAIN       652    756       CADHERIN 4.
FT   DOMAIN       757    858       CADHERIN 5.
FT   DOMAIN       859    961       CADHERIN 6.
FT   DOMAIN       962   1067       CADHERIN 7.
FT   DOMAIN      1068   1169       CADHERIN 8.
FT   DOMAIN      1170   1265       CADHERIN 9.
FT   DOMAIN      1375   1433       EGF-LIKE 1, CALCIUM-BINDING.
FT   DOMAIN      1435   1471       EGF-LIKE 2, CALCIUM-BINDING.
FT   DOMAIN      1475   1514       EGF-LIKE 3, CALCIUM-BINDING.
FT   DOMAIN      1515   1719       LAMININ G-LIKE 1.
FT   DOMAIN      1722   1758       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      1764   1944       LAMININ G-LIKE 2.
FT   DOMAIN      1946   1982       EGF-LIKE 5, CALCIUM-BINDING.
FT   DOMAIN      1983   2020       EGF-LIKE 6, CALCIUM-BINDING.
FT   DOMAIN      2021   2053       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN      2055   2090       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN      2096   2131       LAMININ EGF-LIKE.
FT   DOMAIN      2477   2529       GPS.
FT   DISULFID    1379   1390       BY SIMILARITY.
FT   DISULFID    1384   1421       BY SIMILARITY.
FT   DISULFID    1423   1432       BY SIMILARITY.
FT   DISULFID    1439   1450       BY SIMILARITY.
FT   DISULFID    1444   1459       BY SIMILARITY.
FT   DISULFID    1461   1470       BY SIMILARITY.
FT   DISULFID    1479   1490       BY SIMILARITY.
FT   DISULFID    1484   1500       BY SIMILARITY.
FT   DISULFID    1502   1513       BY SIMILARITY.
FT   DISULFID    1726   1737       BY SIMILARITY.
FT   DISULFID    1731   1746       BY SIMILARITY.
FT   DISULFID    1748   1757       BY SIMILARITY.
FT   DISULFID    1950   1961       BY SIMILARITY.
FT   DISULFID    1955   1970       BY SIMILARITY.
FT   DISULFID    1972   1981       BY SIMILARITY.
FT   DISULFID    1985   1996       BY SIMILARITY.
FT   DISULFID    1990   2008       BY SIMILARITY.
FT   DISULFID    2010   2019       BY SIMILARITY.
FT   DISULFID    2027   2040       BY SIMILARITY.
FT   DISULFID    2042   2052       BY SIMILARITY.
FT   DISULFID    2059   2074       BY SIMILARITY.
FT   DISULFID    2061   2077       BY SIMILARITY.
FT   DISULFID    2079   2089       BY SIMILARITY.
FT   MOD_RES     1963   1963       HYDROXYLATION (POTENTIAL).
FT   CARBOHYD     632    632       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     847    847       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1182   1182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1222   1222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1317   1317       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1327   1327       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1649   1649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1713   1713       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1770   1770       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2053   2053       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2177   2177       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2196   2196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2386   2386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2474   2474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2506   2506       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    2158   2158       G -> GLRGAG (IN REF. 2).
SQ   SEQUENCE    3312 AA;  358251 MW;  BEC208703651A4A5 CRC64;

  Query Match             10.4%;  Score 397.5;  DB 1;  Length 3312;
  Best Local Similarity   23.8%;  Pred. No. 3.8e-17;
  Matches  174;  Conservative   83;  Mismatches  252;  Indels  221;  Gaps   39;

Qy          90 LAPVNGTSQGK----FQGLDLNEELYLGGYPDYGAIPKAGLSS----------------- 128
               |||  |   |     |   :| |:||         : :| |::
Db        1333 LAP-RGAGAGAAGPWFSSEELQEQLY---------VRRAALAARSLLDVLPFDDNVCLRE 1382

Qy         129 ---GFIGCVRELR-------IQGEEIVFHDLNLTAHGISHCP-------------TCRDR 165
                   :: ||  ||       :     :|  :   |     ||              |
Db        1383 PCENYMKCVSVLRFDSSAPFLASASTLFRPIQPIAGLRCRCPPGFTGDFCETELDLCYSN 1442

Qy         166 PCQNGGQCHDSESSSYVCVCPAGFTGSRCE-HSQALHCHPEACGPDATCVNRPDGRGYTC 224
               ||:||| |   |   | |||   |||  ||  ::|  | |  |    || : |:| |: |
Db        1443 PCRNGGACARRE-GGYTCVCRPRFTGEDCELDTEAGRCVPGVCRNGGTCTDAPNG-GFRC 1500

Qy         225 RCHLGRS--GLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLF 282
               :|  | :  | |||       ||     |::    |    | | | : |  :   | |||
Db        1501 QCPAGGAFEGPRCEVAARSFPPS-----SFVMFRGLRQRFH-LTLSLSFATVQQSG-LLF 1553

Qy         283 SGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEP--LALGRWHRVSAERLNK--- 337
                 |:     ||::| :| | :   |  |    |:    |  |: |:|| |     ||
Db        1554 YNGRLNEKHDFLALELVAGQVRLTYSTGESNTVVSPTVPGGLSDGQWHTVHLRYYNKPRT 1613

Qy         338 ---------------------DGSLRVNGGRPV---------LRSSPGKSQGLNLHTLL 366
                                     | :: :  |  :         :::|  ||  |:|   |
Db        1614 DALGGAQGPSKDKVAVLSVDDCDVAVALQFGAEIGNYSCAAAGVQTSSKKS--LDLTGPL 1671

Qy         367 YLGGVEPSVPLS-PATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQ 425
                ||||  |::| : | ::    | ||: :: ::|:|:|:  :|: : |
Db        1672 LLGGV-PNLPENFPVSHKD--FIGCMRDLHIDGRRVDMA-AFVANNG----------TMA 1717

Qy         426 PCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ----GTRCLCL 481
                ||         :  |   |  |                 || : | |        | |
Db        1718 GCQ--------AKLHF---CDSG-----------------PCKNSGFCSERWGSFSCDCP 1749

Qy         482 PGFSGPRCQQGSGHGIAESDWHLEGSG------GNDAPGQYGAYFHDDGFLAFPGHVFSR 535
                || |  ||    |       |  |:|      |:|
Db        1750 VGFGGKDCQLTMAH-----PHHFRGNGTLSWNFGSD------------------------ 1780

Qy         536 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 595
                   ||  : |  ||    |:|:   |: |        :   |  | |      ||| |
Db        1781 MAVSVPWYLGLAFRTRATQGVLM--QVQAGP----HSTLLCQLDRGLLSVTVTRGSGRAS 1834

Qy         596 --LVSEDPINDGEWH--RVTALRE--GRRG----SIQVDGEELVSGRSPGPNVAVNAKGS 645
                 |: :  ::|| ||  |:    |  ||||     : :|        : |  :
Db        1835 HLLLDQVTVSDGRWHDLRLELQEEPGGRRGHHVLMVSLDFSLFQDTMAVGSELQGLKVKQ 1894

Qy         646 VYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAP---PPQPLDLQHRAQA------ 696
               :::|| |     :      |: ||:: : | |   |:|   ||      ||  |
Db        1895 LHVGGLPPG---SAEEAPQGLVGCIQGVWLGSTPSGSPALLPP-----SHRVNAEPGCVV 1946

Qy         697 --GANTRPCP 704
                    : |||
Db        1947 TNACASGPCP 1956

Search completed: March  9, 2004, 17:19:43
Job time : 22.0517 secs
```

```
                   GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:15:08 ; Search time 61.7389 Seconds
                                            (without alignments)
                                            3602.917 Million cell updates/sec

Title:           US-10-006-011A-3
Perfect score:   3825
Sequence:        1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        SPTREMBL_25:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                SUMMARIES
                       %
Result              Query
  No.     Score     Match Length DB  ID                     Description
 -----------------------------------------------------------------------------
     1    901.5     23.6    2026  4  O00468                 O00468 homo sapien
     2      840     22.0     775  4  Q8N7Y0                 Q8n7y0 homo sapien
     3      820     21.4    1009 11   Q8BGP3                 Q8bgp3 mus musculu
     4    812.5     21.2     708  5  Q9NFS9                 Q9nfs9 drosophila
     5    812.5     21.2    1035  5  Q9NEG1                 Q9neg1 drosophila
     6    812.5     21.2    4117  5  Q8IRV9                 Q8irv9 drosophila
     7    812.5     21.2    4179  5  Q9W4Y4                 Q9w4y4 drosophila
     8    812.5     21.2    4223  5  Q8MPN3                 Q8mpn3 drosophila
     9    812.5     21.2    4228  5  Q8IRV8                 Q8irv8 drosophila
    10      703     18.4     488  4  Q96IC1                 Q96ic1 homo sapien
    11    664.5     17.4    1361  5  Q9V7I4                 Q9v7i4 drosophila
    12    664.5     17.4    1361  5  Q9NGV2                 Q9ngv2 drosophila
    13      628     16.4     463  4  Q8NAL2                 Q8nal2 homo sapien
    14    581.5     15.2     822 11   Q8BU56                  Q8bu56 mus musculu
    15    576.5     15.1    1039  5  Q9VQB1                 Q9vqb1 drosophila
    16      511     13.4     925  5  Q9U7E8                 Q9u7e8 caenorhabdi
```

```
    17      479     12.5     350 11  Q8K326                 Q8k326 mus musculu
    18    464.5     12.1     478  5 Q86SD6                Q86sd6 ciona intes
    19    420.5     11.0    1458 11  Q8CJG8                 Q8cjg8 rattus norv
    20    419.5     11.0    1531 11  O88279                 O88279 rattus norv
    21    412.5     10.8    1531 11  Q9WVB5                 Q9wvb5 mus musculu
    22    412.5     10.8    1557 11  Q80TR4                 Q80tr4 mus musculu
    23    411.5     10.8    1530 11  Q9WUG5                 Q9wug5 rattus norv
    24    407.5     10.7    1474 11  Q8CJG9                 Q8cjg9 rattus norv
    25      407     10.6    1534  4 O75093                O75093 homo sapien
    26      407     10.6    1618  4 Q9UIL7                Q9uil7 homo sapien
    27    402.5     10.5    1176 11  Q80Y87                 Q80y87 mus musculu
    28    402.5     10.5    1553 11  Q8CHE6                 Q8che6 mus musculu
    29    394.5     10.3    1523  4 O75094                O75094 homo sapien
    30    393.5     10.3     517  5 Q8MRJ7                Q8mrj7 drosophila
    31      390     10.2    1512 13  Q9DE36                 Q9de36 brachydanio
    32    387.5     10.1    1025 11  Q9Z166                 Q9z166 mus musculu
    33    387.5     10.1    1521 11  Q9R1B9                 Q9r1b9 mus musculu
    34      387     10.1    1523 11  O88280                 O88280 rattus norv
    35    385.5     10.1     950 13  Q90Z44                 Q90z44 gallus gall
    36    385.5     10.1    1392  4 Q8IUE3                Q8iue3 homo sapien
    37      385     10.1    1523 11  Q9WVB4                 Q9wvb4 mus musculu
    38    380.5      9.9     313 13  Q9YHF0                 Q9yhf0 rana pipien
    39    377.5      9.9    1529 13  Q7ZXI2                 Q7zxi2 xenopus lae
    40      375      9.8     738 13  Q90Z45                 Q90z45 gallus gall
    41    370.5      9.7    1100 11  Q8CCT8                 Q8cct8 mus musculu
    42      367      9.6    1515 13  Q9DE37                 Q9de37 brachydanio
    43      363      9.5    1837  5 Q9VCZ9                Q9vcz9 drosophila
    44    362.5      9.5    4307  5 Q19319                Q19319 caenorhabdi
    45    361.5      9.5     792 13  Q90Z43                 Q90z43 gallus gall


                                ALIGNMENTS


RESULT 1
O00468
ID   O00468      PRELIMINARY;      PRT;  2026 AA.
AC   O00468;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   AGRIN precursor (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,
RA   Veerkamp J.H., Monnens L.A.H., van den Heuvel L.P.W.J.;
RL   Eur. J. Biochem. 0:0-0(1998).
RN   [2]
RP   SEQUENCE OF 1-153 FROM N.A.
RX   MEDLINE=96224170; PubMed=8617505;
RA   Lennon G., Auffray C., Polymeropoulos M., Soares M.B.;
RT   "The I.M.A.G.E. Consortium: an integrated molecular analysis of
RT   genomes and their expression.";
RL   Genomics 33:151-152(1996).
RN   [3]
RP   SEQUENCE OF 1-153 FROM N.A.
RA   Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;
RL   J. Cell Biol. 0:0-0(0).
DR   EMBL; AF016903; AAC39776.1; -.
DR   EMBL; U84406; AAB52917.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:329; AGRN.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
```

```
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   InterPro; IPR001455; UPF0033.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 5.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS50024; SEA; 1.
DR   PROSITE; PS01148; UPF0033; 1.
KW   EGF-like domain; Laminin EGF-like domain; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     10       POTENTIAL.
FT   CHAIN        11   2026       AGRIN.
SQ   SEQUENCE   2026 AA;  212881 MW;  4AB0EE710CD4B8EF CRC64;

  Query Match            23.6%;  Score 901.5;  DB 4;  Length 2026;
  Best Local Similarity  32.2%;  Pred. No. 3.2e-58;
  Matches  242;  Conservative   93;  Mismatches  266;  Indels  151;  Gaps   22;

Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              : : ||     |:|||||              |  ||::  |: || : ||| ||| | :
Db       1377 LALEFRALEPQGLLLYNGNA-----------RGKDFLALALLDGRVQLRFDTGSGPAVLT 1425

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD---Y 118
                 |:  | :| : | |   :|:| |    || | |      ||:|: :|::|| |:
Db       1426 SAVPVEPGQWHRLELSRHWRRGTLSVDGETPVLGESPSGTDGLNLDTDLFVGGVPEDQAA 1485

Qy        119 GAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNL------TAHGISHC--PTCRDRPCQNG 170
               |: :  : :|  ||:| | :  : :   :| :         |:  |    |   ||  |
Db       1486 VALERTFVGAGLRGCIRLLDVNNQRL---ELGIGPGAATRGSGVGECGDHPCLPNPCHGG 1542

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
                | :|:   : | || |  |  |   ::   | |  |   | |   |:|  | | |||
Db       1543 APCQNLEAGRFHCQCPPGRVGPTCADEKS-PCQPNPCHGAAPCRVLPEG-GAQCECPLGR 1600

Qy        231 SGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELR--------------LDVEFKPLAP 276
               |  |:     |     |:| :||      :| |||              |:| |    |
Db       1601 EGTFCQ-----TASGQDGSGPFLA-DFNGFSHLELRGLHTFARDLGEKMALEVVFLARGP 1654

Qy        277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLN 336
               |:||::| |:    ||||||:    |||||:|| | ||:|| ||: || | ||| ||
Db       1655 SGLLLYNGQKTDGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNG 1714

Qy        337 KDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSV 396
              : |:|||  |  ||  ||     |||   ||:||      |: |  :|: | | :  ||:
Db       1715 RKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSL 1774

Qy        397 NGKRL----------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQC 443
               |::|          |:| || |           || |    || :||:|:|
Db       1775 GGRQLLTPEHVLRQVDVT-SFAG-----------HPCTRASGHPCLNGASCVP-------- 1815

Qy        444 LCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWH 503
                                  ||            |||  ||||| |::|
Db       1816 ---------------------REAAY--------VCLCPGGFSGPHCEKGL---------- 1837

Qy        504 LEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASG 555
              :| | |            |   ||| |  |   |  | |:         || :||    |
Db       1838 VEKSAG------------DVDTLAFDGRTFVEYLNAVTESEKALQSNHFELSLRTEATQG 1885

Qy        556 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 615
              |:||     |:| :  |:::| : ||||   | |||    | |  |:|   | || | ||
Db       1886 LVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHRE 1941

Qy        616 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNL 673
               | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||::::
Db       1942 QREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDV 2001

Qy        674 VLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |:           || |   |      ||||:
Db       2002 VVGR--------HPLHLLEDAVTKPELRPCPT 2025


RESULT 2
Q8N7Y0
ID   Q8N7Y0      PRELIMINARY;      PRT;   775 AA.
AC   Q8N7Y0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ40230.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK097549; BAC05096.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   775 AA;  84800 MW;  61A81294F0204ED2 CRC64;

  Query Match            22.0%;  Score 840;  DB 4;  Length 775;
  Best Local Similarity  29.6%;  Pred. No. 3.6e-54;
  Matches  214;  Conservative  107;  Mismatches  257;  Indels  144;  Gaps   20;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              :| : || :: ||:||| |:           : : ||:|  ::    :|||: |:|:| |
Db        176 QITLEFRAEAEDGLLLYCGENE---------HGRGDFMSLAIIRRSLQFRFNCGTGVAII 226

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD-YG 119
                 | : || :||| | |    | | : :  || | |||::  :     ||||| |  |
Db        227 VSETKIKLGGWHTVMLYRDGLNGLLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYW 286

Qy        120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN-------LTAHGISHCPT--CRDRPCQNG 170
               :   | : || |||: | :  |  |   |:        |:   :  | : | : |  :|
Db        287 LVRATGTNRGFQGCVQSLAVNGRRI---DMRPWPLGKALSGADVGECSSGICDEASCIHG 343

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   :: ||:|:|| ||
Db        344 GTCTAIKADSYICLCPLGF----------------------------------------- 362

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
```

```
              |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db        363 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 422

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: : |||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db        423 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 477

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        478 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 537

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        538 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 582

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        583 ----LH---CQKAIIEAIEIPQFIG---------------------------RSYLTYDN 608

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  :|:   |||||:|          |||||||:||
Db        609 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 657

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        658 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 717

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        718 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST------DYHISLVEDAVDGKNIN 770

Qy        702 PC 703
               |
Db        771 TC 772

RESULT 3
Q8BGP3
ID   Q8BGP3       PRELIMINARY;      PRT;  1009 AA.
AC   Q8BGP3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical EGF-like domain.
GN   AU040377.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Forelimb, Skin, Testis, and Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK031136; BAC27271.1; -.
DR   EMBL; AK033332; BAC28235.1; -.
DR   EMBL; AK037223; BAC29762.1; -.
DR   EMBL; AK041546; BAC30982.1; -.
DR   MGD; MGI:2146149; AU040377.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
```

```
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   Hypothetical protein.
SQ   SEQUENCE   1009 AA;  109834 MW;  80DD954F255C102B CRC64;

  Query Match             21.4%;  Score 820;  DB 11;  Length 1009;
  Best Local Similarity   29.3%;  Pred. No. 1.6e-52;
  Matches  211;  Conservative  109;  Mismatches  261;  Indels  138;  Gaps   19;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              :: : || :: ||:||| |:           : : ||:|  |:     |||: |:|:| |
Db        410 QVTLEFRAEAEDGLLLYCGESE---------HGRGDFMSLALIRRSLHFRFNCGTGIAII 460

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD-YG 119
                 | : || :||||| |    | | : :  || | |||::  :     ||||| |  |
Db        461 ISETKIKLGAWHTVTLYRDGLNGMLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYW 520

Qy        120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN-------LTAHGISHCPT--CRDRPCQNG 170
               :   | : || |||: | : |::|   |:        |    :  | :  | :  | :|
Db        521 LVRATGTNRGFQGCVQSLSVNGKKI---DMRPWPLGKALNGADVGECSSGICDEASCIHG 577

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   :: ||:|:|| || |  ||              ||  :  |            |
Db        578 GTCAAIKADSYICLCPLGFRGRHCE--------------DAFALTIPQ----------FR 613

Qy        231 SGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS---G 284
                ||                || | |      | |     :: |:| : |||||:|   |
Db        614 ESLR----------------SYAATPWPLEPQHYLSFTEFEITFRPDSGDGVLLYSYDTG 657

Qy        285 GKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVN 344
               |     ||:|: |  ||:|||:: |||  ||||  || ||:|| :   |  |:| |:|:
Db        658 SK-----DFLSINMAAGHVEFRFDCGSGTGVLRSEAPLTLGQWHDLRVSRTAKNGILQVD 712

Qy        345 GGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLT 404
                : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        713 KQKVVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGILHPFSGSIQKIILNDRTIHVK 772

Qy        405 YSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLR 464
              : |  : |:     : ||   || || :| |  |  ::| |  ||:|  |:
Db        773 HDF--TSGVNVENAAHPCVGAPCAHGGSCRPRKE-GYECDCPLGFEGLNCQKAI------ 823

Qy        465 EPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDG 524
                            : :| | |          | :    : ||  |
Db        824 -----------IEAIEIPQFIGRSYLTYDNPNILK---RVSGSRSN-------------A 856

Qy        525 FLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLV 584
              |: |                    :|:   |||||:|          |||||||:|| |:
Db        857 FMRF--------------------KTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALI 894

Qy        585 FRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKG 644
              | | |||| | ::     :|| |||| |:|:|: | | ||     :|:|||    :|  |
Db        895 FSYNLGSGVASIMVNGSFSDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGLMRQLNING 954

Qy        645 SVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 703
              ::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |   |
Db        955 ALYVGGMKEIALHTNRQYLRGLVGCISHFTLST-------DYHISLVEDAVDGKNINTC 1006

RESULT 4
Q9NFS9
ID   Q9NFS9       PRELIMINARY;      PRT;   708 AA.
AC   Q9NFS9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan domain V (Fragment).
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Friedrich M.V.K., Schneider M., Timpl R., Baumgartner S.;
RT   "Perlecan domain V of Drosophila melanogaster: Sequence, recombinant
RT   analysis and tissue expression.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ271462; CAB70094.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   708 AA;  77786 MW;  7B67C6F290642B98 CRC64;

  Query Match             21.2%;  Score 812.5;  DB 5;  Length 708;
  Best Local Similarity   30.2%;  Pred. No. 3.6e-52;
  Matches  219;  Conservative  121;  Mismatches  304;  Indels   81;  Gaps   22;

Qy          3 KITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRH 62
              ::||||:: ||:||:||| |  |          |:|:  |     ||||| |     :|
Db         30 ELTFRPENGDGLLLFNGQTRGSG----------DYIALSLKDRYAEFRFDFGGKPMLVRA 79

Qy         63 PTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIP 122
                ||||  :||| : |    | : | :  ||   :  :   ||| |:||:|| |::  :|
Db         80 EEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVPNWELLP 139

Qy        123 KAGLSS--GFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSES-S 179
                 :|   ||:||:  | :||  :          ||: |  |   |||| | | :|::
Db        140 ADAVSQQVGFVGCISRLTLQGRTVELIREAKYKEGITDCRPCAQGPCQNKGVCLESQTEQ 199

Qy        180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGV 239
              :| |:|  |:||  |   :   | |  ||    | |  :     | | | ||| ||:
Db        200 AYTCICQPGWTGRDCA-IEGTQCTPGVCGA-GRCENTEN--DMECLCPLNRSGDRCQYNE 255

Qy        240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEFKPLA-PDGVLLFSGGKSGPVEDFVS 295
               :   ||:  |:  |    | :|    :: : :  :| :  | |:|::   : |  |:::
Db        256 ILNEHSLNFKGNSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 311

Qy        296 LAMVGGHLEFRYELGSGL--AVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSS 353
              | : ||| |      : |   |:|||||| | || |:   |   :| |||  |      :
Db        312 LVLRGGHAELLINTAARLDPVVVRSAEPLPLNRWTRIEIRRRLGEGILRVGDGPERKAKA 371

Qy        354 PGKSQGLNLHTLLYLGGVEPS-VPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQG 412
              ||  : |:| | ||:|| : | | ::   |::  | ||:  :    | ::|      :
Db        372 PGSDRILSLKTHLYVGGYDRSTVKVNRDVNITKGFDGCISRLYNFQKPVNLLADIKDAAN 431

Qy        413 IGQCYDSSPCERQPCQHGATCMPAGEYE------FQCLCRDGFKGDLCEHEENPCQLREP 466
              |  |             | | |  |: :            |  : :|  :   ||   :|
Db        432 IQSC-------------GETNMIGGDEDSDNEPPVPPPTPDVHENELQPYAMAPC-ASDP 477

Qy        467 CLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHD 522
              | :||:|        | |  ||||  ||:           ||:          : | |
Db        478 CENGGSCSEQEDVAVCSCPFGFSGKHCQE-----------HLQLG--------FNASFRG 518

Qy        523 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGH 582
              ||::      |  :| :   :: :   |:  :||| | | | ||   |:|||:  : ||:
Db        519 DGYVELNRSHFQPALEQSYTSMGIVFTTNKPNGLLFWWGQEAGEEYTGQDFIAAAVVDGY 578

Qy        583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALREGRRGSIQVDGEELVSGRS-PGPNVAV 640
              : :  :|   || : : |  :::|| | | | |:     ::||   | || : |    ::
Db        579 VEYSMRLDGEEAVIRNSDIRVDNGERHIVIAKRDENTAILEVD-RMLHSGETRPTSKKSM 637

Qy        641 NAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANT 700
                 |:|::|||||:   || |:   : ||:  :|:     |      ::|   |  | |
Db        638 KLPGNVFVGGAPDLEVFTGFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 690

Qy        701 RPCPS 705
                ||:
Db        691 NVCPA 695

RESULT 5
Q9NEG1
ID   Q9NEG1      PRELIMINARY;      PRT;  1035 AA.
AC   Q9NEG1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   EG:BACR25B3.11 protein.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murphy L., Harris D., Barrell B.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL138972; CAB72284.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00047; ig; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50835; IG_LIKE; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 1.
DR   PROSITE; PS50068; LDLRA_2; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   EGF-like domain; Immunoglobulin domain.
SQ   SEQUENCE   1035 AA;  114044 MW;  413DE7EDD3AEE711 CRC64;
```

```
  Query Match             21.2%;  Score 812.5;  DB 5;  Length 1035;
  Best Local Similarity   30.2%;  Pred. No. 6e-52;
  Matches  219;  Conservative  121;  Mismatches  304;  Indels   81;  Gaps   22;

Qy          3 KITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRH 62
              ::||||:: ||:||:||| |  |          |:|:  |     ||||| |     :|
Db        370 ELTFRPENGDGLLLFNGQTRGSG----------DYIALSLKDRYAEFRFDFGGKPMLVRA 419

Qy         63 PTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIP 122
                ||||  :||| : |    | : | :  ||   :  :   ||| |:||:|| |::  :|
Db        420 EEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVPNWELLP 479

Qy        123 KAGLSS--GFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSES-S 179
                 :|   ||:||:  | :||. :          ||: |  |   |||| | | :|::
Db        480 ADAVSQQVGFVGCISRLTLQGRTVELIREAKYKEGITDCRPCAQGPCQNKGVCLESQTEQ 539

Qy        180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGV 239
              :| |:|  |:||  |   :   | |  ||    | |  :     | | | ||| ||:
Db        540 AYTCICQPGWTGRDCA-IEGTQCTPGVCGA-GRCENTEN--DMECLCPLNRSGDRCQYNE 595

Qy        240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEFKPLA-PDGVLLFSGGKSGPVEDFVS 295
               :   ||:  |:  |    | :|    :: : :  :| :  | |:|::   : |  |:::
Db        596 ILNEHSLNFKGNSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 651

Qy        296 LAMVGGHLEFRYELGSGL--AVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSS 353
              | : ||| |      : |   |:|||||| | || |:   |   :| |||  |      :
Db        652 LVLRGGHAELLINTAARLDPVVVRSAEPLPLNRWTRIEIRRRLGEGILRVGDGPERKAKA 711

Qy        354 PGKSQGLNLHTLLYLGGVEPS-VPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQG 412
              ||  : |:| | ||:|| : | | ::   |::  | ||:  :    | ::|      :
Db        712 PGSDRILSLKTHLYVGGYDRSTVKVNRDVNITKGFDGCISRLYNFQKPVNLLADIKDAAN 771

Qy        413 IGQCYDSSPCERQPCQHGATCMPAGEYE------FQCLCRDGFKGDLCEHEENPCQLREP 466
              |  |             | | |  |: :            |  : :|  :   ||   :|
Db        772 IQSC-------------GETNMIGGDEDSDNEPPVPPPTPDVHENELQPYAMAPC-ASDP 817

Qy        467 CLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHD 522
              | :||:|        | |  ||||  ||:           ||:          : | |
Db        818 CENGGSCSEQEDVAVCSCPFGFSGKHCQE-----------HLQLG--------FNASFRG 858

Qy        523 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGH 582
              ||::      |  :| :   :: :   |:  :||| | | | ||   |:|||:  : ||:
Db        859 DGYVELNRSHFQPALEQSYTSMGIVFTTNKPNGLLFWWGQEAGEEYTGQDFIAAAVVDGY 918

Qy        583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALREGRRGSIQVDGEELVSGRS-PGPNVAV 640
              : :  :|   || : : |  :::|| | | | |:     ::||   | || : |    ::
Db        919 VEYSMRLDGEEAVIRNSDIRVDNGERHIVIAKRDENTAILEVD-RMLHSGETRPTSKKSM 977

Qy        641 NAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANT 700
                 |:|::|||||:   || |:   : ||:  :|:     |      ::|   |  | |
Db        978 KLPGNVFVGGAPDLEVFTGFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 1030

Qy        701 RPCPS 705
                ||:
Db       1031 NVCPA 1035


RESULT 6
Q8IRV9
ID   Q8IRV9      PRELIMINARY;      PRT;  4117 AA.
AC   Q8IRV9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7981-PC.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
```

```
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003424; AAN09077.1; -.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00047; ig; 12.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 23.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 12.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 23.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 11.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 20.
DR   PROSITE; PS50068; LDLRA_2; 23.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
SQ   SEQUENCE   4117 AA;  454841 MW;  BB47056B7E4D6106 CRC64;

  Query Match             21.2%;  Score 812.5;  DB 5;  Length 4117;
  Best Local Similarity   30.2%;  Pred. No. 3.9e-51;
  Matches  219;  Conservative  121;  Mismatches  304;  Indels   81;  Gaps   22;

Qy          3 KITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRH 62
              ::||||:: ||:||:||| |  |          |:|:  |     ||||| |     :|
Db       3439 ELTFRPENGDGLLLFNGQTRGSG----------DYIALSLKDRYAEFRFDFGGKPMLVRA 3488

Qy         63 PTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIP 122
                ||||  :||| : |    | : | :  ||   :  :   ||| |:||:|| |::  :|
Db       3489 EEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVPNWELLP 3548

Qy        123 KAGLSS--GFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSES-S 179
                 :|   ||:||:  | :||  :          ||: |  |   |||| | | :|::
Db       3549 ADAVSQQVGFVGCISRLTLQGRTVELIREAKYKEGITDCRPCAQGPCQNKGVCLESQTEQ 3608

Qy        180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGV 239
              :| |:|  |:||  |   :   | |  ||    | |  :     | | | ||| ||:
Db       3609 AYTCICQPGWTGRDCA-IEGTQCTPGVCGA-GRCENTEN--DMECLCPLNRSGDRCQYNE 3664

Qy        240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEFKPLA-PDGVLLFSGGKSGPVEDFVS 295
               :   ||:  |:  |    | :|    :: : :  :| :  | |:|::   : |  |:::
Db       3665 ILNEHSLNFKGNSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 3720

Qy        296 LAMVGGHLEFRYELGSGL--AVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSS 353
              | : ||| |      : |   |:|||||| | || |:   |   :| |||  |      :
Db       3721 LVLRGGHAELLINTAARLDPVVVRSAEPLPLNRWTRIEIRRRLGEGILRVGDGPERKAKA 3780

Qy        354 PGKSQGLNLHTLLYLGGVEPS-VPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQG 412
              ||  : |:| | ||:|| : | | ::   |::  | ||:  :    | ::|      :
Db     , 3781 PGSDRILSLKTHLYVGGYDRSTVKVNRDVNITKGFDGCISRLYNFQKPVNLLADIKDAAN 3840

Qy        413 IGQCYDSSPCERQPCQHGATCMPAGEYE------FQCLCRDGFKGDLCEHEENPCQLREP 466
              |  |             | | |  |: :            |  : :|  :   ||   :|
Db       3841 IQSC-------------GETNMIGGDEDSDNEPPVPPPTPDVHENELQPYAMAPC-ASDP 3886

Qy        467 CLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHD 522
              | :||:|        | |  ||||  ||:           ||:          : | |
Db       3887 CENGGSCSEQEDVAVCSCPFGFSGKHCQE-----------HLQLG--------FNASFRG 3927

Qy        523 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGH 582
              ||::      |  :| :   :: :   |:  :||| | | | ||   |:|||:  : ||:
Db       3928 DGYVELNRSHFQPALEQSYTSMGIVFTTNKPNGLLFWWGQEAGEEYTGQDFIAAAVVDGY 3987

Qy        583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALREGRRGSIQVDGEELVSGRS-PGPNVAV 640
              : :  :|   || : : |  :::|| | | | |:    ::||    | || : |    ::
Db       3988 VEYSMRLDGEEAVIRNSDIRVDNGERHIVIAKRDENTAILEVD-RMLHSGETRPTSKKSM 4046

Qy        641 NAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANT 700
                 |:|::|||||:   || |:   : ||:  :|:     |      ::|   |  | |
Db       4047 KLPGNVFVGGAPDLEVFTGFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 4099

Qy        701 RPCPS 705
                ||:
Db       4100 NVCPA 4104

RESULT 7
Q9W4Y4
ID   Q9W4Y4      PRELIMINARY;      PRT;  4179 AA.
AC   Q9W4Y4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7981 protein.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

```
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003424; AAF45786.3; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00047; ig; 12.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 23.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00180; EGF_Lam; 9.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 12.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 23.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 11.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 20.
DR   PROSITE; PS50068; LDLRA_2; 23.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Immunoglobulin domain; Laminin EGF-like domain.
SQ   SEQUENCE   4179 AA;  461782 MW;  47804277D0914E63 CRC64;

  Query Match             21.2%;  Score 812.5;  DB 5;  Length 4179;
  Best Local Similarity   30.2%;  Pred. No. 4e-51;
  Matches  219;  Conservative  121;  Mismatches  304;  Indels   81;  Gaps   22;

Qy          3 KITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRH 62
              ::||||:: ||:||:||| |  |          |:|:  |     ||||| |     :|
Db       3501 ELTFRPENGDGLLLFNGQTRGSG----------DYIALSLKDRYAEFRFDFGGKPMLVRA 3550

Qy         63 PTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIP 122
                ||||  :||| : |    | : | :  ||   :  :   ||| |:||:|| |::  :|
Db       3551 EEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVPNWELLP 3610

Qy        123 KAGLSS--GFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSES-S 179
                 :|   ||:||:  | :||  :          ||: |  |   |||| | | :|::
Db       3611 ADAVSQQVGFVGCISRLTLQGRTVELIREAKYKEGITDCRPCAQGPCQNKGVCLESQTEQ 3670

Qy        180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGV 239
              :| |:|  |:||  |   :   | |  ||    | |  :     | | | ||| ||:
Db       3671 AYTCICQPGWTGRDCA-IEGTQCTPGVCGA-GRCENTEN--DMECLCPLNRSGDRCQYNE 3726

Qy        240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEFKPLA-PDGVLLFSGGKSGPVEDFVS 295
               :   ||:  |:  |    | :|    :: : :  :| :  | |:|::   : |  |:::
Db       3727 ILNEHSLNFKGNSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 3782

Qy        296 LAMVGGHLEFRYELGSGL--AVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSS 353
              | : ||| |      : |   |:|||||| | || |:   |   :| |||  |      :
Db       3783 LVLRGGHAELLINTAARLDPVVVRSAEPLPLNRWTRIEIRRRLGEGILRVGDGPERKAKA 3842

Qy        354 PGKSQGLNLHTLLYLGGVEPS-VPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQG 412
              ||  : |:| | ||:|| : | | ::   |::  | ||:  :    | ::|      :
Db       3843 PGSDRILSLKTHLYVGGYDRSTVKVNRDVNITKGFDGCISRLYNFQKPVNLLADIKDAAN 3902

Qy        413 IGQCYDSSPCERQPCQHGATCMPAGEYE------FQCLCRDGFKGDLCEHEENPCQLREP 466
              |  |             | | |  |: :            |  : :|  :   ||   :|
Db       3903 IQSC-------------GETNMIGGDEDSDNEPPVPPPTPDVHENELQPYAMAPC-ASDP 3948

Qy        467 CLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHD 522
              | :||:|        | |  ||||  ||:           ||:          : | |
Db       3949 CENGGSCSEQEDVAVCSCPFGFSGKHCQE-----------HLQLG--------FNASFRG 3989

Qy        523 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGH 582
              ||::      |  :| :   :: :   |:  :||| | | | ||   |:|||:  : ||:
Db       3990 DGYVELNRSHFQPALEQSYTSMGIVFTTNKPNGLLFWWGQEAGEEYTGQDFIAAAVVDGY 4049

Qy        583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALREGRRGSIQVDGEELVSGRS-PGPNVAV 640
              : :  :|   || : : |  :::|| | | | |:     ::||   | || : |    ::
Db       4050 VEYSMRLDGEEAVIRNSDIRVDNGERHIVIAKRDENTAILEVD-RMLHSGETRPTSKKSM 4108

Qy        641 NAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANT 700
                 |:|::|||||:   || |:   : ||:  :|:     |      ::|   |  | |
Db       4109 KLPGNVFVGGAPDLEVFTGFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 4161

Qy        701 RPCPS 705
                ||:
Db       4162 NVCPA 4166
```

```
RESULT 8
Q8MPN3
ID   Q8MPN3      PRELIMINARY;      PRT;  4223 AA.
AC   Q8MPN3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan (CG7981-PD).
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Voigt A.;
RT   "Perlecan participates in proliferation activation of quiescent
RT   Drosophila neuroblasts.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ487018; CAD31650.1; -.
DR   EMBL; AE003424; AAN09079.1; -.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00047; ig; 13.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 23.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 8.
DR   SMART; SM00409; IG; 13.
DR   SMART; SM00408; IGc2; 13.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 23.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 12.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 20.
DR   PROSITE; PS50068; LDLRA_2; 23.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   4223 AA;  466919 MW;  31D7C2B5C9B1D2E0 CRC64;

  Query Match             21.2%;  Score 812.5;  DB 5;  Length 4223;
  Best Local Similarity   30.2%;  Pred. No. 4e-51;
  Matches  219;  Conservative  121;  Mismatches  304;  Indels   81;  Gaps   22;
```

```
Qy          3 KITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRH 62
              ::||||:: ||:||:||| |  |          |:|:  |     ||||| |     :|
Db       3545 ELTFRPENGDGLLLFNGQTRGSG----------DYIALSLKDRYAEFRFDFGGKPMLVRA 3594

Qy         63 PTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIP 122
                ||||  :||| : |    | : | :  ||   :  :   ||| |:||:|| |::  :|
Db       3595 EEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVPNWELLP 3654

Qy        123 KAGLSS--GFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSES-S 179
                 :|   ||:||:  | :||  :          ||: |  |   |||| | | :|::
Db       3655 ADAVSQQVGFVGCISRLTLQGRTVELIREAKYKEGITDCRPCAQGPCQNKGVCLESQTEQ 3714

Qy        180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGV 239
              :| |:|  |:||  |   :   | |  ||    | |  :     | | | ||| ||:
Db       3715 AYTCICQPGWTGRDCA-IEGTQCTPGVCGA-GRCENTEN--DMECLCPLNRSGDRCQYNE 3770

Qy        240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEFKPLA-PDGVLLFSGGKSGPVEDFVS 295
               :   ||:  |:  |    | :|    :: : :  :| :  | |:|::   : |  |:::
Db       3771 ILNEHSLNFKGNSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 3826

Qy        296 LAMVGGHLEFRYELGSGL--AVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSS 353
              | : ||| |      : |   |:|||||| | || |:   |   :| |||  |      :
Db       3827 LVLRGGHAELLINTAARLDPVVVRSAEPLPLNRWTRIEIRRRLGEGILRVGDGPERKAKA 3886

Qy        354 PGKSQGLNLHTLLYLGGVEPS-VPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQG 412
              ||  : |:| | ||:|| : | | ::   |::  | ||:  :    | ::|      :
Db       3887 PGSDRILSLKTHLYVGGYDRSTVKVNRDVNITKGFDGCISRLYNFQKPVNLLADIKDAAN 3946

Qy        413 IGQCYDSSPCERQPCQHGATCMPAGEYE------FQCLCRDGFKGDLCEHEENPCQLREP 466
              |  |             | | |  |: :            |  : :|  :   ||   :|
Db       3947 IQSC-------------GETNMIGGDEDSDNEPPVPPPTPDVHENELQPYAMAPC-ASDP 3992

Qy        467 CLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHD 522
              | :||:|        | |  ||||  ||:           ||:          : | |
Db       3993 CENGGSCSEQEDVAVCSCPFGFSGKHCQE-----------HLQLG--------FNASFRG 4033

Qy        523 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGH 582
              ||::      |  :| :   :: :   |:  :||| | | | ||   |:|||:  : ||:
Db       4034 DGYVELNRSHFQPALEQSYTSMGIVFTTNKPNGLLFWWGQEAGEEYTGQDFIAAAVVDGY 4093

Qy        583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALREGRRGSIQVDGEELVSGRS-PGPNVAV 640
              : :  :|   || : : |  :::|| | | | |:     ::||   | || : |    ::
Db       4094 VEYSMRLDGEEAVIRNSDIRVDNGERHIVIAKRDENTAILEVD-RMLHSGETRPTSKKSM 4152

Qy        641 NAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANT 700
                 |:|::|||||:   || |:   : ||:  :|:     |      ::|   |  | |
Db       4153 KLPGNVFVGGAPDLEVFTGFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 4205

Qy        701 RPCPS 705
                ||:
Db       4206 NVCPA 4210
```

```
RESULT 9
Q8IRV8
ID   Q8IRV8      PRELIMINARY;      PRT;  4228 AA.
AC   Q8IRV8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7981-PE.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wamg Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003424; AAN09078.1; -.
DR   FlyBase; FBgn0001402; trol.
```

```
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00047; ig; 12.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 24.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 12.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 24.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 11.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 21.
DR   PROSITE; PS50068; LDLRA_2; 24.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
SQ   SEQUENCE   4228 AA;  466934 MW;  0F630AA0BBF4CD33 CRC64;

  Query Match             21.2%;  Score 812.5;  DB 5;  Length 4228;
  Best Local Similarity   30.2%;  Pred. No. 4e-51;
  Matches  219;  Conservative  121;  Mismatches  304;  Indels   81;  Gaps   22;

Qy            3 KITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRH 62
                ::||||:: ||:||:||| |  |          |:|:  |     ||||| |     :|
Db         3550 ELTFRPENGDGLLLFNGQTRGSG----------DYIALSLKDRYAEFRFDFGGKPMLVRA 3599

Qy           63 PTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIP 122
                  ||||  :||| : |    | : | :  ||   :  :   ||| |:||:|| |::  :|
Db         3600 EEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVPNWELLP 3659

Qy          123 KAGLSS--GFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSES-S 179
                   :|   ||:||:  | :||  :          ||: |  |   |||| | | :|::
Db         3660 ADAVSQQVGFVGCISRLTLQGRTVELIREAKYKEGITDCRPCAQGPCQNKGVCLESQTEQ 3719

Qy          180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGV 239
                :| |:|  |:||  |   :   | |  ||    | |  :     | | | ||| ||:
Db         3720 AYTCICQPGWTGRDCA-IEGTQCTPGVCGA-GRCENTEN--DMECLCPLNRSGDRCQYNE 3775

Qy          240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEFKPLA-PDGVLLFSGGKSGPVEDFVS 295
                 :   ||:  |:  |    | :|    :: : :  :| :  | |:|::   : |  |:::
Db         3776 ILNEHSLNFKGNSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 3831

Qy          296 LAMVGGHLEFRYELGSGL--AVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSS 353
                | : ||| |      : |   |:|||||| | || |:   |   :| |||  |      :
Db         3832 LVLRGGHAELLINTAARLDPVVVRSAEPLPLNRWTRIEIRRRLGEGILRVGDGPERKAKA 3891

Qy          354 PGKSQGLNLHTLLYLGGVEPS-VPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQG 412
                ||  : |:| | ||:|| : | | ::   |::  | ||:  :    | ::|      :
Db         3892 PGSDRILSLKTHLYVGGYDRSTVKVNRDVNITKGFDGCISRLYNFQKPVNLLADIKDAAN 3951

Qy          413 IGQCYDSSPCERQPCQHGATCMPAGEYE------FQCLCRDGFKGDLCEHEENPCQLREP 466
                |  |             | | |  |: :            |  : :|  :   ||   :|
Db         3952 IQSC-------------GETNMIGGDEDSDNEPPVPPPTPDVHENELQPYAMAPC-ASDP 3997

Qy          467 CLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHD 522
                | :||:|        | |  ||||  ||:           ||:          : | |
Db         3998 CENGGSCSEQEDVAVCSCPFGFSGKHCQE-----------HLQLG--------FNASFRG 4038

Qy          523 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGH 582
                ||::      |  :| :   :: :   |:  :||| | | | ||   |:|||:  : ||:
Db         4039 DGYVELNRSHFQPALEQSYTSMGIVFTTNKPNGLLFWWGQEAGEEYTGQDFIAAAVVDGY 4098

Qy          583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALREGRRGSIQVDGEELVSGRS-PGPNVAV 640
                : :  :|   || : : |  :::|| | | | |:     ::||   | || : |    ::
Db         4099 VEYSMRLDGEEAVIRNSDIRVDNGERHIVIAKRDENTAILEVD-RMLHSGETRPTSKKSM 4157

Qy          641 NAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANT 700
                   |:|::|||||:   || |:   : ||:  :|:     |      ::|   |  | |
Db         4158 KLPGNVFVGGAPDLEVFTGFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 4210

Qy          701 RPCPS 705
                  ||:
Db         4211 NVCPA 4215

RESULT 10
Q96IC1
ID   Q96IC1      PRELIMINARY;      PRT;   488 AA.
AC   Q96IC1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007649; AAH07649.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   488 AA;  51764 MW;  A732B99CC680E772 CRC64;

  Query Match             18.4%;  Score 703;  DB 4;  Length 488;
  Best Local Similarity   33.2%;  Pred. No. 3.4e-44;
  Matches  191;  Conservative   63;  Mismatches  196;  Indels  126;  Gaps   17;

Qy          167 CQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRC 226
                |  |  | : |:  : | || |  |  |   ::  | |  |   | |   |:| |  | |
Db            1 CHGGAPCQNLEAGRFHCQCPPGRVGPTCADEKS-PCQPNPCHGAAPCRVLPEG-GAQCEC 58

Qy          227 HLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELR--------------LDVEFK 272
                 ||| |  |:     |     |:| :||      :| |||              |:| |
Db           59 PLGREGTFCQ-----TASGQDGSGPFLA-DFNGFSHLELRGLHTFARDLGEKMALEVVFL 112

Qy          273 PLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSA 332
                   | |:||::| |:    ||||||:    |||||:|| | ||:|| ||: || | |||
Db          113 ARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSL 172
```

```
Qy          333 ERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVG 392
                ||  : |:|||  |  ||  ||     |||   ||:||      |: |  :|: | | :
Db          173 ERNGRKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQ 232

Qy          393 EVSVNGKRL----------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEY 439
                 ||: |::|          |:| || |           || |    || :||:|:|
Db          233 LVSLGGRQLLTPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP---- 277

Qy          440 EFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAE 499
                                        ||            |||  ||||| |::|
Db          278 ------------------------REAAY--------VCLCPGGFSGPHCEKGL------ 299

Qy          500 SDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTS 551
                    :| | |            |   |||  |  |   |  | |:         || :||
Db          300 ----VEKSAG------------DVDTLAFDGRTFVEYLNAVTESEKALQSNHFELSLRTE 343

Qy          552 TASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVT 611
                   ||:||     |:| :  |:::| : ||||   |  |||    | |  |:|   |  ||
Db          344 ATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVV 399

Qy          612 ALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGC 669
                | || |  ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||
Db          400 AHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGC 459

Qy          670 VKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                ::::|:           || |   |      ||||:
Db          460 LRDVVVGR--------HPLHLLEDAVTKPELRPCPT 487

RESULT 11
Q9V7I4
ID   Q9V7I4      PRELIMINARY;      PRT;  1361 AA.
AC   Q9V7I4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG8403 protein.
GN   SP2353 OR CG8403.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003808; AAF58071.2; -.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0034070; SP2353.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 4.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain.
SQ   SEQUENCE   1361 AA;  150298 MW;  4C71EDA81A12ABED CRC64;

  Query Match             17.4%;  Score 664.5;  DB 5;  Length 1361;
  Best Local Similarity   22.0%;  Pred. No. 1e-40;
  Matches  221;  Conservative  111;  Mismatches  295;  Indels  379;  Gaps   23;

Qy            1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
                :::| |||:| ||::| :|::          :   ||::  |  |  || || |||: ::
Db          424 QLRIEFRPESFDGIILLSGER---------DDLTGDFMALLLNKGFVEFWFDCGSGVGSV 474
```

```
Qy        61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDY-G 119
             |    : |  :::| : |      |::     | | | | |  :   | ::|||  :  |
Db       475 RSRETILLNEWNSVIIYRHRWDAWLVLNHGTKVQGRSNGLFSRITFREPVFLGGIGNITG 534

Qy       120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAH---------GISHCPT--CRDRPCQ 168
                :  |: || ||:|  |    |   ||   | |          |  | |  |   |||
Db       535 LAKRLPLAEGFAGCIR--RFVANE---HDYKFTEHPLGDVINGFDIQDCSTDKCVRYPCQ 589

Qy       169 NGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
             :||:|  |:  : :|:|| || |  ||
Db       590 HGGKCLPSDQGA-ICLCPIGFVGDLCE--------------------------------- 615

Qy       229 GRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSG 288
                 :| :    :  |: :|:      |   :    | | |  ||   ||::|:|| :
Db       616 ----IRMD----LQVPAFNGSSFLRYAPLGDSALIWLELKVTLKPEQADGLILYSGPEHR 667

Qy       289 PVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRP 348
                ||::| :  | :|| ::|||| |::||   |:||:|| :   |  :   |:|:  :
Db       668 --GDFIALYLNDGFVEFAFDLGSGPALVRSEHSLSLGQWHTIKISRTARLAVLKVDKHQE 725

Qy       349 VLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFL 408
             ||  |      |:|   |::|||     |         | ||:  : :||  | :    |
Db       726 VLTISSNGFWHLSLDQNLFVGGVNHVDRLPLDLKYKPFFVGCIQRIDINGHSLGIVAEAL 785

Qy       409 GSQGIGQCYDSSPCERQPCQHGATCMPAGE-YEFQCLCRDGFKGDLCEH--EENPCQLRE 465
             |   || |     |  :||   | |:|  | ||    ||     : |    |  | || |
Db       786 GGSNIGNC--PHACVARPCGPLAECVPQMESYE----CRCSIHNERCNKAAEVPPEQLPE 839

Qy       466 PCLHGGTCQGTR------------------------------------------------ 477
               ||      |:
Db       840 LALHKSKVLETKDNGEAAKKVSGLAHKKHSKKHRNLHKPTPATSTTTSTTSTTTTTTEAP 899

Qy       478 --------------------------------------------------------CLCL 481

Db       900 SERTEEATAGALSNEEIEDDIIFRLVQQQQQQKELKKQHQQTTTTAPATSTTSSGPSKAK 959

Qy       482 PGFSG------------------------------------------------------- 486
             |  ||
Db       960 PRLSGKHHASKHEHHLKPNAAFTRKLSRLPTHYESFQTNPDSDILTFEDNNDWVTSLQQQ 1019

Qy       487 ----------------------PRCQQG-------------------------------- 492
                                   ||
Db      1020 EYGDAMAASQVPLAFEDASPGTPRSSDNNEDDENAFVFDESLFDASDGTEEYQRKQLAQD 1079

Qy       493 -----------SGHGIAESDWHLEGS----GGNDAPGQYGAYFHDDGFL----------- 526
                        | |  ||  :  :||      |:   ||   ::||  |
Db      1080 MKRIMSNSNAHSSHKKAEVQFPPQGSQEVGTANEDTSQYSDDYNDDELLTPVMQGGEEVK 1139

Qy       527 ----------------------------------------AFPG--------------HV 532
                                                      | |               |
Db      1140 LEQHTSSTPQTHTDWSLLKKFDLSAEHQSQVQGVRKNFGACFAGSDSYFHYNDADTMSQV 1199

Qy       533 FSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSG 592
              | |:      : | ::| : :|::|| | : |   :  |::|||:: |:| ||| ||||
Db      1200 ISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLSLGIEQGYLHFRYDLGSG 1253

Qy       593 EARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGA 651
             |  :  :   ::|| |||| |:|  : | ::||| : |: |:||    :|    :|:||
Db      1254 EVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAPGKLRQLNTDTGLYVGGM 1313

Qy       652 PDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAG 697
             |||   |  |: ||| ||:  :||         |  |  :|  : |
Db      1314 PDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNVETG 1359

RESULT 12
Q9NGV2
ID   Q9NGV2      PRELIMINARY;      PRT;  1361 AA.
AC   Q9NGV2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SP2353.
GN   SP2353 OR CG8403.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Serano T.L., Pendleton J.D., Rubin G.M.;
RT   "A reverse genetic screen for genes involved in Drosophila
RT   development.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF239610; AAF63502.1; -.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0034070; SP2353.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain.
SQ   SEQUENCE   1361 AA;  150355 MW;  206D9F9BF9EDFE47 CRC64;

  Query Match             17.4%;  Score 664.5;  DB 5;  Length 1361;
  Best Local Similarity   22.0%;  Pred. No. 1e-40;
  Matches  221;  Conservative  111;  Mismatches  295;  Indels  379;  Gaps   23;

Qy         1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
             :::| |||:| ||::| :|::          :   ||::  |  |  || || |||: ::
Db       424 QLRIEFRPESFDGIILLSGER---------DDLTGDFMALLLNKGFVEFWFDCGSGVGSV 474

Qy        61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDY-G 119
             |    : |  :::| : |      |::     | | | | |  :   | ::|||  :  |
Db       475 RSRETILLNEWNSVIIYRHRWDAWLVLNHGTKVQGRSNGLFSRITFREPVFLGGIGNITG 534

Qy       120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAH---------GISHCPT--CRDRPCQ 168
                :  |: || ||:|  |    |   ||   | |          |  | |  |   |||
Db       535 LAKRLPLAEGFAGCIR--RFVANE---HDYKFTEHPLGDVINGFDIQDCSTDKCVRYPCQ 589

Qy       169 NGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
             :||:|  |:  : :|:|| || |  ||
Db       590 HGGKCLPSDQGA-ICLCPIGFVGDLCE--------------------------------- 615

Qy       229 GRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSG 288
                 :| :    :  |: :|:      |   :    | | |  ||   ||::|:|| :
Db       616 ----IRMD----LQVPAFNGSSFLRYAPLGDSALIWLELKVTLKPEQADGLILYSGPEHR 667

Qy       289 PVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRP 348
                ||::| :  | :|| ::|||| |::||   |:||:|| :   |  :   |:|:  :
Db       668 --GDFIALYLNDGFVEFAFDLGSGPALVRSEHSLSLGQWHTIKISRTARLAVLKVDKHQE 725

Qy       349 VLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFL 408
             ||  |      |:|   |::|||     |         | ||:  : :||  | :    |
Db       726 VLTISSNGFWHLSLDQNLFVGGVNHVDRLPLDLKYKPFFVGCIQRIDINGHSLGIVAEAL 785

Qy       409 GSQGIGQCYDSSPCERQPCQHGATCMPAGE-YEFQCLCRDGFKGDLCEH--EENPCQLRE 465
             |   || |     |  :||   | |:|  | ||    ||     : |    |  | || |
Db       786 GGSNIGNC--PHACVARPCGPLAECVPQMESYE----CRCSIHNERCNKAAEVPPEQLPE 839

Qy       466 PCLHGGTCQGTR------------------------------------------------ 477
               ||      |:
Db       840 LALHKSKVLETKDNGEAAKKVSGLAHKKHSKKHRNLHKPTPATSTTTSTTSTTTTTTEAP 899

Qy       478 --------------------------------------------------------CLCL 481

Db       900 SERTEEATAGALSNEEIEDDIIFRLVQQQQQQKELKKQHQQTTTTAPATSTTSSGPSKAK 959
```

```
Qy         482 PGFSG------------------------------------------------------- 486
               |  ||
Db         960 PRLSGKHHASKHEHHLKPNAAFTRKLSRLPTHYESFQTNPDSDILTFEDNNDWVTSLQQQ 1019

Qy         487 ----------------------PRCQQG-------------------------------- 492
                                     ||
Db        1020 EYGDAMAASQVPLAFEDASPGTPRSSDNNEDDENAFVFDESLFDASDGTEEYQRKQLAQD 1079

Qy         493 -----------SGHGIAESDWHLEGS----GGNDAPGQYGAYFHDDGFL----------- 526
                          | |  ||  :  :||      |:   ||   ::||  |
Db        1080 MKRIMSNSNAHSSHKKAEVQFPPQGSQEVGTANEDTSQYSDDYNDDELLTPVMQGGEEVK 1139

Qy         527 ----------------------------------------AFPG--------------HV 532
                                                        | |               |
Db        1140 LEQHTSSTPQTHTDWSLLKKFDLSAEHQSQVQGVRKNFGACFAGSDSYFHYNDADTMSQV 1199

Qy         533 FSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSG 592
                | |:      : | ::| : :|::|| | : |   :  |::|||:: |:| ||| ||||
Db        1200 ISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLSLGIEQGYLHFRYDLGSG 1253

Qy         593 EARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGA 651
               |  :  :   ::|| |||| |:|  : | ::||| : |: |:||    :|    :|:||
Db        1254 EVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAPGKLRQLNTDTGLYVGGM 1313

Qy         652 PDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAG 697
               |||   |  |: ||| ||:  :||         |  |  :|  : |
Db        1314 PDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNVETG 1359

RESULT 13
Q8NAL2
ID   Q8NAL2        PRELIMINARY;      PRT;   463 AA.
AC   Q8NAL2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ35160.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK092479; BAC03900.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   463 AA;  50635 MW;  F688BF2714D5D0EC CRC64;

  Query Match             16.4%;  Score 628;  DB 4;  Length 463;
  Best Local Similarity   28.7%;  Pred. No. 1.3e-38;
  Matches  162;  Conservative   80;  Mismatches  198;  Indels  124;  Gaps   16;
```

```
Qy         151 LTAHGISHCPT--CRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACG 208
               |:   :  | :  | :  | :|| |   ::  ||:|:|| ||
Db          10 LSGADVGECSSGICDEASCIHGGTCTAIKADSYICLCPLGF------------------- 50

Qy         209 PDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL-- 265
                                      |  ||:  |:| |       || | |      | |
Db          51 ----------------------KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSF 88

Qy         266 -RLDVEFKPLAPDGVLLFS---GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEP 321
                  :: |:| : |||||:|   | |     ||:|: : |||:|||:: |||  |||| :|
Db          89 MEFEITFRPDSGDGVLLYSYDTGSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDP 143

Qy         322 LALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPAT 381
               | || || :   |  |:| |:|:  : |   : |    :  :| :::|||     :   :
Db         144 LTLGNWHELRVSRTAKNGILQVDKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNS 203

Qy         382 NMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEF 441
                :   | | : :: :| : : : : |  : |:     : || | || || :| |  |  :
Db         204 GVLKPFSGSIQKIILNDRTIHVKHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GY 260

Qy         442 QCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAE 499
                | |  ||:|                ||   ||      : :| | |
Db         261 DCDCPLGFEG----------------LH---CQKAIIEAIEIPQFIG------------- 288

Qy         500 SDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLW 559
                               |  | : |      |   |||       : :  :|:   |||||
Db         289 --------------RSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLW 325

Qy         560 QGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRG 619
               :|          |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: |
Db         326 RG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSG 383

Qy         620 SIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSAR 679
                | ||     :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db         384 KITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-- 441

Qy         680 PGAPPPQPLDLQHRAQAGANTRPC 703
                       : |   |  | |   |
Db         442 -----DYHISLVEDAVDGKNINTC 460

RESULT 14
Q8BU56
ID   Q8BU56        PRELIMINARY;      PRT;   822 AA.
AC   Q8BU56;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical EGF-like domain.
GN   AU040377.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Oviduct;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK087659; BAC39956.1; -.
DR   MGD; MGI:2146149; AU040377.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
```

```
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00054; laminin_G; 2.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   822 AA;  89669 MW;  225A01D56D8DE691 CRC64;

  Query Match             15.2%;  Score 581.5;  DB 11;  Length 822;
  Best Local Similarity   30.3%;  Pred. No. 8.7e-35;
  Matches  139;  Conservative   70;  Mismatches  184;  Indels   65;  Gaps   10;

Qy        200 LHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALT 259
              : |    |  |: |||     |  | |:||: |  | | : :  |   |  ||:    |
Db        345 MSCDETLCSADSFCVNDYAWGGSRCHCNLGKGGEACSEDIFIQYPQFFG-HSYVTFEPLK 403

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSA 319
              |::   :: :||:  | ||:||: |       |||:|||::   | ||:  |:|:|:: |
Db        404 NSYQAFQVTLEFRAEAEDGLLLYCGESEHGRGDFMSLALIRRSLHFRFNCGTGIAIIISE 463

Qy        320 EPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSP 379
                : || || |:  |   :| |::| | ||   | |:   :   | |||||   :  |
Db        464 TKIKLGAWHTVTLYRDGLNGMLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYWLVR 523

Qy        380 ATNMSAHFRGCVGEVSVNGKRLDLTYSFLG----SQGIGQCYDSSPCERQPCQHGATCMP 435
              ||  :  |:|||  :|||||::|:    ||       :|:|  |  |:   | || ||
Db        524 ATGTNRGFQGCVQSLSVNGKKIDMRPWPLGKALNGADVGEC-SSGICDEASCIHGGTCAA 582

Qy        436 AGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGH 495
                   : |||  ||:|    |                |:    | :| |      : |
Db        583 IKADSYICLCPLGFRG---RH----------------CEDAFALTIPQF------RESLR 617

Qy        496 GIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 555
                | : | ||                       | |  | :        |:  |  :  |
Db        618 SYAATPWPLE-----------------------PQHYLSFT------EFEITFRPDSGDG 648

Qy        556 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 615
              :||:   : |     |||:|: :  ||: ||:  |||   | || |:  |:|| :   |
Db        649 VLLY-SYDTG----SKDFLSINMAAGHVEFRFDCGSGTGVLRSEAPLTLGQWHDLRVSRT 703

Qy        616 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD 653
               : | :||| :::| | : |    :     ::||| |:
Db        704 AKNGILQVDKQKVVEGMAEGGFTQIKCNTDIFIGGVPN 741

RESULT 15
Q9VQB1
ID   Q9VQB1      PRELIMINARY;      PRT;  1039 AA.
AC   Q9VQB1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7245 protein.
GN   CG7245.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003583; AAF51266.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0031416; CG7245.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 6.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   EGF-like domain.
SQ   SEQUENCE   1039 AA;  114305 MW;  8864EC2A0F345590 CRC64;

  Query Match             15.1%;  Score 576.5;  DB 5;  Length 1039;
  Best Local Similarity   27.2%;  Pred. No. 2.8e-34;
  Matches  200;  Conservative   95;  Mismatches  298;  Indels  143;  Gaps   31;

Qy         70 HFHTVTL--LRSLT--QGSLIVGDLAPVNGTSQGKFQGLD---------LNEELYLGGYP 116
              |  |: |   |: |    ||:| |   ::|      : |          ||  |:||| |
Db          2 HCSTIRLDFSRNYTHCNASLLVNDTLAMSGDQPTWLKLLPPRLHTPEAILNTWLHLGGAP 61

Qy        117 D--YGAI---PKAGLSSGFIGCVRELRIQG----------EEIVFHDLN-----LTAHGI 156
                  | |   | |   ||| ||:  ||| |          |: :  :|      |   ||
Db         62 QAPIGLIIELPPAQSGSGFTGCLHTLRINGQAREIFGQPDEKDISKELGGRSDALDGFGI 121

Qy        157 SHCPT--CRDRPCQNGGQC----------HDSESSSYVCVCPAGFTGSRCEHSQALHCHP 204
              : | :  |   ||:||  |          :  ::  : | || |: |  || |    |
Db        122 TECGSLACLSSPCRNGAACIKIETNDLDENGEKAEKWKCKCPTGYMGPTCEISV---CED 178
```

```
Qy          205 EACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGA----GSYLALPALTN 260
                  |    |||  | | || | | ||: |  ||  : |  || ||:     |::|
Db          179 NPCQYGGTCVQFP-GSGYLCLCPLGKHGHYCEHNLEVALPSFSGSVNGLSSFVAYTVPIP 237

Qy          261 THHELRLDVEFKP--LAPDGVLLFSG--GKSGPVEDFVSLAMVGGHLEFRYELGSGLAVL 316
                  : | |  :  |  ::   :| | |  |      | :::: : |::   : ||:|   :
Db          238 LEYSLELSFKILPQTMSQISLLAFFGQSGYHDEKSDHLAVSFIQGYIMLTWNLGAGPRRI 297

Qy          317 RSAEPL-----ALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGV 371
                 : :|:     |    : :   |: :   | |:|   :   |||    :::  :|||||
Db          298 FTQKPIDFRLDAPRVPYEIKVGRIGRQAWLSVDGKFNITGRSPGSGSRMDVLPILYLGGH 357

Qy          372 E--------PSVPLSPATNMSAHFRGCVGEVSVNGKR----LDLTYSFLGSQGIGQCYDS 419
                |          :||       : |:||: :| :   :    |  |    | :|:|||  :
Db          358 EIANFNTLPHDLPL------HSGFQGCIYDVQLKAGQVTVPLQETRGVRG-RGVGQC-GT 409

Qy          420 SPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPC-QLREPCLHGGTC----Q 474
                  | |  |||   |:  |   | |:|::|: | ||    |||      |    ||
Db          410 RECHRHACQHDGACLQHGA-TFTCICQEGWYGPLCAQPTNPCDSFNNKCYEDATCVPLVN 468

Qy          475 GTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGS------------GGNDAPG----QYGA 518
                |  | |  | :|  |::      : ||  | |              | |  |    |  :
Db          469 GYECDCPVGRTGKNCEEVIR---SLSDVSLTGRRSYLAVRWPYLYDGGDKLGAKRSQMVS 525

Qy          519 YFHDDGFLAFP-------GHVFSRSLPEVPET------------------------IELE 547
                | :    |  |        |   : | || :                         || :
Db          526 YRNFTKKLMPPKPITTPSSHFVMKLLNEVEKQRSFSPVPLMGSKSFEEHHRVQFFFIEFQ 585

Qy          548 VRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV-SEDPINDGE 606
                :|  :  ||||: |       :   |:|| || | : ||    |    :| |   :  ||
Db          586 LRPLSERGLLLYFGTLNNNQDKKIGFVSLSLQGGVVEFRISGPSNHVTVVRSVRMLAIGE 645

Qy          607 WHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSS-- 664
                ||::   : ||  :: |:|    |  :|   | |     :||||  ||: |     |
Db          646 WHKIKMAQRGRWLTLWVEG-SASSALAPSAEVLVEPDSLLYIGGLKDVSKLPHNAISGFP 704

Qy          665 -GITGCVKNLVLHSAR 679
                    |||: ||:   |
Db          705 IPFRGCVRGLVVSGTR 720


Search completed: March  9, 2004, 17:21:42
Job time : 67.739 secs
```

```
                         GenCore version 5.1.6
                 Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:19:48 ; Search time 85.8465 Seconds
                                             (without alignments)
                                             2320.373 Million cell updates/sec

Title:           US-10-006-011A-3
Perfect score:   705
Sequence:        1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        1586107 seqs, 282547505 residues

Word size :       0

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        A_Geneseq_29Jan04:*
                 1:  geneseqp1980s:*
                 2:  geneseqp1990s:*
                 3:  geneseqp2000s:*
                 4:  geneseqp2001s:*
                 5:  geneseqp2002s:*
                 6:  geneseqp2003as:*
                 7:  geneseqp2003bs:*
                 8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                     %
Result               Query
  No.     Score      Match Length DB  ID                  Description
-------------------------------------------------------------------------------
    1      705  100.0   4391  6  AAE34390             Aae34390 Human per
    2      443   62.8   4393  4  AAB31889             Aab31889 Amino aci
    3      383   54.3   4436  4  ABG23265             Abg23265 Novel hum
    4      195   27.7    195  4  AAB31890             Aab31890 Amino aci
    5       18    2.6     18  4  ABB55911             Abb55911 Vascular
    6       14    2.0     14  4  ABB55908             Abb55908 Vascular
    7       14    2.0     14  4  ABB55905             Abb55905 Vascular
    8       10    1.4     10  4  ABB55910             Abb55910 Vascular
    9        9    1.3     15  4  ABB55909             Abb55909 Vascular
   10        9    1.3    159  6  ABM65732             Abm65732 Propionib
   11        9    1.3   2531  7  ADE63713             Ade63713 Rat Prote
   12        9    1.3   2531  7  ADE63705             Ade63705 Rat Prote
   13        9    1.3   2531  7  ADE63709             Ade63709 Rat Prote
   14        9    1.3   2531  7  ADE63701             Ade63701 Rat Prote
   15        9    1.3   3319  4  ABB70376             Abb70376 Drosophil
   16        8    1.1      8  4  ABB55907             Abb55907 Vascular
   17        8    1.1     14  4  ABB55906             Abb55906 Vascular
   18        8    1.1     36  5  AAE13652             Aae13652 Mouse cyc
   19        8    1.1     47  5  AAE13653             Aae13653 Human e-s
   20        8    1.1     54  4  AAU40544             Aau40544 Propionib
   21        8    1.1     54  6  ABM37063             Abm37063 Propionib
   22        8    1.1     54  7  ABO23342             Abo23342 Human sec
   23        8    1.1     55  2  AAY01436             Aay01436 Secreted
   24        8    1.1     66  5  AAE13647             Aae13647 Human fib
   25        8    1.1     77  4  AAU63351             Aau63351 Propionib
   26        8    1.1     77  6  ABM59870             Abm59870 Propionib
   27        8    1.1     80  4  AAM16081             Aam16081 Peptide #
   28        8    1.1     80  4  ABB35070             Abb35070 Peptide #
   29        8    1.1     80  4  AAM28574             Aam28574 Peptide #
   30        8    1.1     80  4  ABB29890             Abb29890 Peptide #
   31        8    1.1     80  4  ABB20486             Abb20486 Protein #
   32        8    1.1     80  4  AAM68258             Aam68258 Human bon
   33        8    1.1     80  4  AAM55885             Aam55885 Human bra
   34        8    1.1     80  4  ABG49912             Abg49912 Human liv
   35        8    1.1     80  4  AAM03808             Aam03808 Peptide #
   36        8    1.1     80  5  ABG37793             Abg37793 Human pep
   37        8    1.1     82  5  AAE13646             Aae13646 Human coa
   38        8    1.1     83  5  AAE13645             Aae13645 Pig facto
   39        8    1.1     88  5  AAE13648             Aae13648 Human coa
   40        8    1.1    110  5  ABB53124             Abb53124 Human ORF
   41        8    1.1    123  5  AAE13650             Aae13650 Human ins
   42        8    1.1    131  5  ABB90171             Abb90171 Human pol
   43        8    1.1    143  5  AAE13643             Aae13643 Mouse lam
   44        8    1.1    144  5  ABB53123             Abb53123 Human ORF
   45        8    1.1    144  5  ABB53141             Abb53141 Human ORF

                              ALIGNMENTS

RESULT 1
AAE34390
ID   AAE34390 standard; protein; 4391 AA.
XX
AC   AAE34390;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Human perlecan protein.
XX
KW   Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX
OS   Homo sapiens.
XX
PN   WO200295415-A2.
XX
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-EP005612.
XX
PR   23-MAY-2001; 2001GB-00012626.
XX
PA   (OSTE-) OSTEOMETER BIO TECH AS.
XX
PI   Christgau S,  Henriksen DB,  Cloos PAC;
XX
DR   WPI; 2003-140389/13.
XX
PT   An assay for the diagnosis or assessment of the severity of
PT   osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
PT   or optically inverted protein in a sample.
XX
PS   Disclosure; Page 46-67; 106pp; English.
XX
CC   The invention relates to an assay for the diagnosis or assessment of the
CC   severity of osteoarthritis or rheumatoid arthritis. The assay involves
CC   measuring (in a biological sample) the amount or presence of an
CC   isomerised or optically inverted protein or one or more isomerised or
CC   optically inverted fragments from proteins such as perlecan, biglycan,
CC   decorin, fibrillin-1 or protocadherin. The assay is useful for the
CC   diagnosis or assessment of the severity of osteoarthritis or rheumatoid
CC   arthritis. The present sequence is human perlecan protein
XX
SQ   Sequence 4391 AA;

  Query Match             100.0%;  Score 705;  DB 6;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  705;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3687 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3746

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3747 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3806

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3807 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3866

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3867 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3926

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3927 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3987 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4046

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4047 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4106

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4107 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4166

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4167 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4226

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4227 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4286

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4287 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 4346

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4347 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391

RESULT 2
AAB31889
ID   AAB31889 standard; protein; 4393 AA.
XX
AC   AAB31889;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a human protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 138-152; 209pp; French.
XX
CC   The present sequence represents a human protein, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 4393 AA;

  Query Match             62.8%;  Score 443;  DB 4;  Length 4393;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches  643;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3689 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3748

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3749 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3808

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3809 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3868

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3869 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3928

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3929 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3988

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       3989 GHLEFRYELGSGLAVLRTAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4048

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4049 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4108

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       4109 PCERQPCQHGATCMPAGEYEFQCLCRDGIKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4168

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4169 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4228
```

```
Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4229 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4288

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGS 645
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4289 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGS 4333

RESULT 3
ABG23265
ID   ABG23265 standard; protein; 4436 AA.
XX
AC   ABG23265;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23256.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87452.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 53624; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4436 AA;

  Query Match             54.3%;  Score 383;  DB 4;  Length 4436;
  Best Local Similarity   99.7%;  Pred. No. 0;
```

```
  Matches  583;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3722 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3781

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3782 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3841

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3842 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3901

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3902 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3961

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3962 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 4021

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       4022 GHLEFRYELGSGLAVLRTAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4081

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       4082 NLHTLLYLGGVEPSVPLSPATNMGAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4141

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4142 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4201

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4202 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4261

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVF 585
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4262 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVF 4306


RESULT 4
AAB31890
ID   AAB31890 standard; protein; 195 AA.
XX
AC   AAB31890;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of the C-terminal of the human perlecan protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
```

```
DR   WPI; 2001-159475/16.
DR   N-PSDB; AAF54728.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 152-153; 209pp; French.
XX
CC   The present sequence represents a human polypeptide, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 195 AA;

  Query Match             27.7%;  Score 195;  DB 4;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-183;
  Matches  195;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        511 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 60

Qy        571 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 120

Qy        631 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 180

Qy        691 QHRAQAGANTRPCPS 705
              |||||||||||||||
Db        181 QHRAQAGANTRPCPS 195


RESULT 5
ABB55911
ID   ABB55911 standard; peptide; 18 AA.
XX
AC   ABB55911;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 111.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 18 AA;

  Query Match             2.6%;  Score 18;  DB 4;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-10;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        644 GSVYIGGAPDVATLTGGR 661
              ||||||||||||||||||
Db          1 GSVYIGGAPDVATLTGGR 18


RESULT 6
ABB55908
ID   ABB55908 standard; peptide; 14 AA.
XX
AC   ABB55908;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 108.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
```

```
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 14 AA;

  Query Match             2.0%;  Score 14;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        596 LVSEDPINDGEWHR 609
              ||||||||||||||
Db          1 LVSEDPINDGEWHR 14

RESULT 7
ABB55905
ID   ABB55905 standard; peptide; 14 AA.
XX
AC   ABB55905;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 105.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 14 AA;

  Query Match             2.0%;  Score 14;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        619 GSIQVDGEELVSGR 632
              ||||||||||||||
Db          1 GSIQVDGEELVSGR 14

RESULT 8
ABB55910
ID   ABB55910 standard; peptide; 10 AA.
XX
AC   ABB55910;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 110.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
```

```
XX
SQ   Sequence 10 AA;

  Query Match             1.4%;  Score 10;  DB 4;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.041;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        662 FSSGITGCVK 671
              ||||||||||
Db          1 FSSGITGCVK 10

RESULT 9
ABB55909
ID   ABB55909 standard; peptide; 15 AA.
XX
AC   ABB55909;
XX
DT   15-FEB-2002  (first entry)
XX
DE   Vascular dementia-associated protein isoform (VPI) 109.
XX
KW   Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
KW   diagnosis; prognosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200169261-A2.
XX
PD   20-SEP-2001.
XX
PF   14-MAR-2001; 2001WO-GB001106.
XX
PR   15-MAR-2000; 2000GB-00006285.
PR   24-NOV-2000; 2000GB-00028734.
PR   28-NOV-2000; 2000US-00724391.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Herath HMAC,  Parekh RB,  Rohlff C;
XX
DR   WPI; 2001-557937/62.
XX
PT   Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT   determining stage of VD and monitoring the effect of VD therapy,
PT   comprises analyzing body fluid by 2-dimensional electrophoresis for
PT   features correlated with VD.
XX
PS   Claim 6; Page 32; 151pp; English.
XX
CC   The invention relates to screening, diagnosis or prognosis of Vascular
CC   Dementia (VD) in a subject comprising analysing body fluid from the
CC   subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
CC   features containing at least one chosen feature whose relative abundance
CC   correlates with the presence, absence, stage or severity of VD or
CC   predicts the onset or course of VD, especially detecting in a sample of
CC   cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
CC   protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
CC   specification. Detecting VD-associated features and VPI is useful for the
CC   screening, diagnosis or prognosis of VD, for determining the stage or
CC   severity of VD, for identifying a subject at risk of VD or for monitoring
CC   the effect of therapy administered to a subject having VD. Nucleic acids
CC   encoding a VPI or inhibiting the function of a VPI are useful for the
CC   treatment of VD and for gene therapy
XX
SQ   Sequence 15 AA;

  Query Match             1.3%;  Score 9;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.58;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        587 YQLGSGEAR 595
              |||||||||
Db          1 YQLGSGEAR 9

RESULT 10
ABM65732
ID   ABM65732 standard; protein; 159 AA.
XX
AC   ABM65732;
XX
DT   20-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes immunogenic polypeptide #30408.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine; immunogenic.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Claim 7; SEQ ID NO 30408; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a specifically claimed P. acnes polypeptide which is
CC   thought to contain an immunogenic region. Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 159 AA;

  Query Match             1.3%;  Score 9;  DB 6;  Length 159;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        548 VRTSTASGL 556
              |||||||||
Db         31 VRTSTASGL 39


RESULT 11
ADE63713
ID   ADE63713 standard; protein; 2531 AA.
XX
AC   ADE63713;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9657.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2531 AA;
```

```
  Query Match             1.3%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity  100.0%;  Pred. No. 74;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 12
ADE63705
ID   ADE63705 standard; protein; 2531 AA.
XX
AC   ADE63705;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9649.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
```

```
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2531 AA;

  Query Match             1.3%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 74;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 13
ADE63709
ID   ADE63709 standard; protein; 2531 AA.
XX
AC   ADE63709;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9653.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
```

```
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2531 AA;

  Query Match             1.3%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 74;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 14
ADE63701
ID   ADE63701 standard; protein; 2531 AA.
XX
AC   ADE63701;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rat Protein CAA40667, SEQ ID NO 9645.
XX
KW   Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW   chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS   Rattus norvegicus.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; CAA40667.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
```

```
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a rat protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2531 AA;

  Query Match             1.3%;  Score 9;  DB 7;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 74;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 15
ABB70376
ID   ABB70376 standard; protein; 3319 AA.
XX
AC   ABB70376;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 37920.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL14479.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 37920; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3319 AA;

  Query Match             1.3%;  Score 9;  DB 4;  Length 3319;
  Best Local Similarity   100.0%;  Pred. No. 96;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 NDGEWHRVT 611
              |||||||||
Db       3033 NDGEWHRVT 3041

Search completed: March  9, 2004, 17:27:53
Job time : 88.8465 secs
```

```
                         GenCore version 5.1.6
                Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:25:24 ; Search time 27.0475 Seconds
                                            (without alignments)
                                            1345.642 Million cell updates/sec

Title:          US-10-006-011A-3
Perfect score:  705
Sequence:       1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       389414 seqs, 51625971 residues

Word size :       0

Total number of hits satisfying chosen parameters:       389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        Issued_Patents_AA:*
               1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
               2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
               3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
               4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
               5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
               6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                   SUMMARIES
                        %
Result                Query
  No.     Score  Match Length DB  ID                          Description
-----------------------------------------------------------------------------
    1         9    1.3   1940  2  US-08-644-271-30           Sequence 30, Appl
    2         9    1.3   1940  4  US-09-077-955-34           Sequence 34, Appl
    3         8    1.1    241  4  US-09-107-532A-5434        Sequence 5434, Ap
    4         8    1.1    427  4  US-09-252-991A-19300       Sequence 19300, A
    5         8    1.1    579  4  US-09-540-236-2071         Sequence 2071, Ap
    6         8    1.1   1015  1  US-08-537-210A-1           Sequence 1, Appli
    7         8    1.1   1015  3  US-09-113-825-1            Sequence 1, Appli
    8         8    1.1   1130  2  US-08-460-309-2            Sequence 2, Appli
    9         8    1.1   1130  2  US-08-125-077-2            Sequence 2, Appli
   10         8    1.1   1130  6  5444158-2                 Patent No. 5444158
   11         8    1.1   1257  1  US-08-340-428B-49          Sequence 49, Appl
   12         8    1.1   2321  4  US-09-230-652-2            Sequence 2, Appli
   13         8    1.1   2471  1  US-08-185-432-16           Sequence 16, Appl
   14         8    1.1   2471  1  US-08-083-590A-19          Sequence 19, Appl
   15         8    1.1   2471  3  US-08-532-384-19           Sequence 19, Appl
   16         8    1.1   2471  4  US-08-899-232-1            Sequence 1, Appli
   17         8    1.1   3084  4  US-09-562-702A-12          Sequence 12, Appl
   18         8    1.1   3088  4  US-09-562-702A-8           Sequence 8, Appli
   19         8    1.1   3089  4  US-09-562-702A-4           Sequence 4, Appli
   20         8    1.1   3106  4  US-09-562-702A-10          Sequence 10, Appl
   21         8    1.1   3110  4  US-09-562-702A-2           Sequence 2, Appli
   22         8    1.1   3110  4  US-09-562-702A-6           Sequence 6, Appli
   23         8    1.1   3110  4  US-09-561-709B-7           Sequence 7, Appli
   24         8    1.1   3111  2  US-08-460-309-4            Sequence 4, Appli
   25         8    1.1   3111  2  US-08-125-077-4            Sequence 4, Appli
   26         7    1.0     77  4  US-09-252-991A-18061       Sequence 18061, A
   27         7    1.0    100  4  US-09-249-697A-3           Sequence 3, Appli
```

```
   28         7    1.0    100  4  US-09-363-316B-3           Sequence 3, Appli
   29         7    1.0    108  4  US-09-621-976-4170         Sequence 4170, Ap
   30         7    1.0    146  4  US-09-543-681A-7317        Sequence 7317, Ap
   31         7    1.0    162  4  US-09-107-532A-5488        Sequence 5488, Ap
   32         7    1.0    165  4  US-09-252-991A-27759       Sequence 27759, A
   33         7    1.0    181  4  US-09-252-991A-23483       Sequence 23483, A
   34         7    1.0    186  4  US-09-252-991A-21008       Sequence 21008, A
   35         7    1.0    186  4  US-09-107-532A-6672        Sequence 6672, Ap
   36         7    1.0    201  4  US-09-252-991A-28786       Sequence 28786, A
   37         7    1.0    216  1  US-08-315-695-20           Sequence 20, Appl
   38         7    1.0    237  4  US-09-252-991A-28116       Sequence 28116, A
   39         7    1.0    240  4  US-09-570-856B-16          Sequence 16, Appl
   40         7    1.0    244  4  US-09-252-991A-18465       Sequence 18465, A
   41         7    1.0    248  2  US-08-755-559-1            Sequence 1, Appli
   42         7    1.0    248  3  US-09-210-474-1            Sequence 1, Appli
   43         7    1.0    248  4  US-09-539-774-1            Sequence 1, Appli
   44         7    1.0    273  4  US-09-252-991A-28324       Sequence 28324, A
   45         7    1.0    290  4  US-09-252-991A-27099       Sequence 27099, A

                                  ALIGNMENTS


RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1940 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Rat Agrin
;      LOCATION:  1...1940
;      OTHER INFORMATION:
US-08-644-271-30

  Query Match              1.3%;  Score 9;  DB 2;  Length 1940;
```

```
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1228 PCLHGGTCQ 1236


RESULT 2
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-077-955-34

  Query Match             1.3%;  Score 9;  DB 4;  Length 1940;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1228 PCLHGGTCQ 1236


RESULT 3
US-09-107-532A-5434
; Sequence 5434, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 5434:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 241 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...241
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5434:
US-09-107-532A-5434

  Query Match             1.1%;  Score 8;  DB 4;  Length 241;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        144 IVFHDLNL 151
              ||||||||
Db         28 IVFHDLNL 35


RESULT 4
US-09-252-991A-19300
; Sequence 19300, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 19300
;   LENGTH: 427
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19300

  Query Match             1.1%;  Score 8;  DB 4;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         76 LLRSLTQG 83
              ||||||||
Db        247 LLRSLTQG 254


RESULT 5
US-09-540-236-2071
; Sequence 2071, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2071
;   LENGTH: 579
```

```
;   TYPE: PRT
;   ORGANISM: M.catarrhalis
US-09-540-236-2071

  Query Match              1.1%;  Score 8;  DB 4;  Length 579;
  Best Local Similarity  100.0%;  Pred. No. 56;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        319 AEPLALGR 326
              ||||||||
Db        320 AEPLALGR 327


RESULT 6
US-08-537-210A-1
; Sequence 1, Application US/08537210A
; Patent No. 5780300
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Fortini, Mark
;    APPLICANT:  Matsuno, Kenji
;    TITLE OF INVENTION:  MANIPULATION OF NON-TERMINALLY
;    TITLE OF INVENTION:  DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/537,210A
;      FILING DATE:  29-SEP-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  hum N (Human No. 5780300ch 2)
;      LOCATION:  1155...2169
;      OTHER INFORMATION:  Highly conserved ankyrin repeat
;      OTHER INFORMATION:  region of No. 5780300ch
US-08-537-210A-1

  Query Match              1.1%;  Score 8;  DB 1;  Length 1015;
  Best Local Similarity  100.0%;  Pred. No. 94;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 PCQHGATC 433
              ||||||||
Db          3 PCQHGATC 10


RESULT 7
US-09-113-825-1
; Sequence 1, Application US/09113825
; Patent No. 6149902
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Fortini, Mark
;    APPLICANT:  Matsuno, Kenji
;    TITLE OF INVENTION:  MANIPULATION OF NON-TERMINALLY
;    TITLE OF INVENTION:  DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/113,825
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/537,210
;      FILING DATE:  29-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1015 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  hum N (Human No. 6149902ch 2)
;      LOCATION:  1155...2169
;      OTHER INFORMATION:  Highly conserved ankyrin repeat
;      OTHER INFORMATION:  region of No. 6149902ch
US-09-113-825-1

  Query Match              1.1%;  Score 8;  DB 3;  Length 1015;
  Best Local Similarity  100.0%;  Pred. No. 94;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 PCQHGATC 433
              ||||||||
Db          3 PCQHGATC 10


RESULT 8
US-08-460-309-2
; Sequence 2, Application US/08460309
; Patent No. 5837496
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
```

```
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,309
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,077
;      FILING DATE:  22-SEP-1993
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-460-309-2

  Query Match              1.1%;  Score 8;  DB 2;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        543 TIELEVRT 550
              ||||||||
Db        808 TIELEVRT 815


RESULT 9
US-08-125-077-2
; Sequence 2, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,077
;      FILING DATE:  22-SEP-1993
```

```
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1130 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-125-077-2

  Query Match              1.1%;  Score 8;  DB 2;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        543 TIELEVRT 550
              ||||||||
Db        808 TIELEVRT 815


RESULT 10
5444158-2
;Patent No. 5444158
;    APPLICANT: ENGVALL,. EVA;SANES, JOSHUA
;    TITLE OF INVENTION: MEROSIN, NUCLEIC ACIDS ENCODING,
;FRAGMENTS AND USES THEREOF
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/87,642
;      FILING DATE: 08-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 587,689
;      FILING DATE: 24-SEP-1990
;      APPLICATION NUMBER: 472,319
;      FILING DATE: 30-JAN-1990
;SEQ ID NO:2:
;      LENGTH: 1130
5444158-2

  Query Match              1.1%;  Score 8;  DB 6;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        543 TIELEVRT 550
              ||||||||
Db        808 TIELEVRT 815


RESULT 11
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
;  GENERAL INFORMATION:
;    APPLICANT:  MARGOLIS, Richard U.
;    APPLICANT:  RAUCH, Uwe
;    APPLICANT:  MARGOLIS, Renee K.
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES FOR A
;    TITLE OF INVENTION:  NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
;    NUMBER OF SEQUENCES:  49
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/340,428B
;      FILING DATE:  14 No. 5648465ember 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/922,911
;      FILING DATE:  03 August 1992
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Browdy, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  Margolis=1A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1257 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-340-428B-49

  Query Match             1.1%;  Score 8;  DB 1;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTC 473
              ||||||||
Db        957 PCLHGGTC 964


RESULT 12
US-09-230-652-2
; Sequence 2, Application US/09230652A
; Patent No. 6537775
; GENERAL INFORMATION:
;  APPLICANT: Tournier-Lasserve, Elisabeth
;  APPLICANT: Joutel, Anne
;  APPLICANT: Bousser, Marie-Germaine
;  APPLICANT: Bach, Jean-Francois
;  TITLE OF INVENTION: GENE INVOLVED IN CADASIL, METHOD OF DIAGNOSIS AND
;  TITLE OF INVENTION: THERAPEUTIC APPLICATION
;  FILE REFERENCE: 03715.0048-00000
;  CURRENT APPLICATION NUMBER: US/09/230,652A
;  CURRENT FILING DATE: 1999-05-17
;  EARLIER APPLICATION NUMBER: FR 96 09733
;  EARLIER FILING DATE: 1996-08-01
;  EARLIER APPLICATION NUMBER: FR 97 04680
;  EARLIER FILING DATE: 1997-04-16
;  EARLIER APPLICATION NUMBER: PCT/FR97/01433
;  EARLIER FILING DATE: 1997-07-31
;  NUMBER OF SEQ ID NOS: 163
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 2321
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human ADNc No. 6537775ch 3
```

```
US-09-230-652-2

  Query Match             1.1%;  Score 8;  DB 4;  Length 2321;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        483 GFSGPRCQ 490
              ||||||||
Db       1237 GFSGPRCQ 1244


RESULT 13
US-08-185-432-16
; Sequence 16, Application US/08185432
; Patent No. 5750652
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Busseau, Isabelle
;    APPLICANT:  Diederich, Robert J.
;    APPLICANT:  Xu, Tian
;    APPLICANT:  Matsuno, Kenji
;    TITLE OF INVENTION:  DELTEX PROTEINS, NUCLEIC ACIDS, AND
;    TITLE OF INVENTION:  ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/185,432
;      FILING DATE:  21-JAN-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-185-432-16

  Query Match             1.1%;  Score 8;  DB 1;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 PCQHGATC 433
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 14
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 5786158ch Proteins And
```

```
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/083,590A
;      FILING DATE:  25-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-083-590A-19

  Query Match              1.1%;  Score 8;  DB 1;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 PCQHGATC 433
              ||||||||
Db       1157 PCQHGATC 1164


RESULT 15
US-08-532-384-19
; Sequence 19, Application US/08532384
; Patent No. 6083904
;  GENERAL INFORMATION:
;    APPLICANT:  Artavanis-Tsakonas, S. et al.
;    TITLE OF INVENTION:  Therapeutic And Diagnostic Methods
;    TITLE OF INVENTION:  And Compositions Based On No. 6083904ch Proteins And
;    TITLE OF INVENTION:  Nucleic Acids
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/532,384
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/083,590
;      FILING DATE:  25-JUN-1993
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7326-015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 790-9090
;      TELEFAX:  212 8698864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2471 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-532-384-19

  Query Match              1.1%;  Score 8;  DB 3;  Length 2471;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 PCQHGATC 433
              ||||||||
Db       1157 PCQHGATC 1164

Search completed: March  9, 2004, 17:32:24
Job time : 28.0475 secs
```

```
                      GenCore version 5.1.6
             Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:27:59 ; Search time 51.7431 Seconds
                                            (without alignments)
                                            2876.963 Million cell updates/sec

Title:           US-10-006-011A-3
Perfect score:   705
Sequence:        1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        809742 seqs, 211153259 residues

Word size :        0

Total number of hits satisfying chosen parameters:      809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        Published_Applications_AA:*
                1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result           Query
  No.   Score    Match Length DB  ID                       Description
 -----------------------------------------------------------------------
    1       9     1.3   1940  13  US-10-016-283-34          Sequence 34, Appl
    2       9     1.3   2531  15  US-10-190-115-29          Sequence 29, Appl
    3       9     1.3   2531  15  US-10-369-072-29          Sequence 29, Appl
    4       8     1.1     54  10  US-09-776-724A-136        Sequence 136, App
    5       8     1.1     80   9  US-09-864-761-35784       Sequence 35784, A
    6       8     1.1    131  15  US-10-264-237-2547        Sequence 2547, Ap
    7       8     1.1    264  11  US-09-833-245-2022        Sequence 2022, Ap
    8       8     1.1    264  15  US-10-266-829-75          Sequence 75, Appl
    9       8     1.1    299  14  US-10-219-220-251         Sequence 251, App
   10       8     1.1    309  14  US-10-122-706-31          Sequence 31, Appl
   11       8     1.1    383  11  US-09-833-245-770         Sequence 770, App
   12       8     1.1    383  14  US-10-245-103-80          Sequence 80, Appl
   13       8     1.1    383  14  US-10-245-107-80          Sequence 80, Appl
   14       8     1.1    383  14  US-10-245-143-80          Sequence 80, Appl
   15       8     1.1    383  14  US-10-245-771-80          Sequence 80, Appl
   16       8     1.1    383  14  US-10-245-851-80          Sequence 80, Appl
   17       8     1.1    383  14  US-10-245-883-80          Sequence 80, Appl
   18       8     1.1    383  14  US-10-237-535-80          Sequence 80, Appl
   19       8     1.1    383  14  US-10-238-183-80          Sequence 80, Appl
   20       8     1.1    383  14  US-10-238-283-80          Sequence 80, Appl
   21       8     1.1    383  14  US-10-238-370-80          Sequence 80, Appl
   22       8     1.1    383  14  US-10-245-055-80          Sequence 80, Appl
   23       8     1.1    383  14  US-10-245-147-80          Sequence 80, Appl
   24       8     1.1    383  14  US-10-245-730-80          Sequence 80, Appl
   25       8     1.1    383  14  US-10-245-739-80          Sequence 80, Appl
   26       8     1.1    383  14  US-10-246-210-80          Sequence 80, Appl
   27       8     1.1    383  14  US-10-239-196-80          Sequence 80, Appl
   28       8     1.1    383  14  US-10-243-024-80          Sequence 80, Appl
   29       8     1.1    383  14  US-10-243-409-80          Sequence 80, Appl
   30       8     1.1    383  14  US-10-245-621-80          Sequence 80, Appl
   31       8     1.1    383  14  US-10-245-880-80          Sequence 80, Appl
   32       8     1.1    383  14  US-10-245-033-80          Sequence 80, Appl
   33       8     1.1    383  14  US-10-243-095-80          Sequence 80, Appl
   34       8     1.1    383  14  US-10-245-185-80          Sequence 80, Appl
   35       8     1.1    383  14  US-10-245-427-80          Sequence 80, Appl
   36       8     1.1    383  14  US-10-245-473-80          Sequence 80, Appl
   37       8     1.1    383  14  US-10-245-770-80          Sequence 80, Appl
   38       8     1.1    383  14  US-10-245-877-80          Sequence 80, Appl
   39       8     1.1    383  14  US-10-246-976-80          Sequence 80, Appl
   40       8     1.1    383  14  US-10-243-320-80          Sequence 80, Appl
   41       8     1.1    383  14  US-10-242-743-80          Sequence 80, Appl
   42       8     1.1    383  14  US-10-242-845-80          Sequence 80, Appl
   43       8     1.1    383  14  US-10-237-636-80          Sequence 80, Appl
   44       8     1.1    383  14  US-10-238-325-80          Sequence 80, Appl
   45       8     1.1    383  14  US-10-238-346-80          Sequence 80, Appl


                              ALIGNMENTS


RESULT 1
US-10-016-283-34
; Sequence 34, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/10/016,283
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US/09/077,955A
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: PCT/US96/20696
;  PRIOR FILING DATE: 1996-12-13
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-10-016-283-34

  Query Match             1.3%;  Score 9;  DB 13;  Length 1940;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1228 PCLHGGTCQ 1236


RESULT 2
US-10-190-115-29
; Sequence 29, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook, John P. II
;  APPLICANT:  Boldog, Ferenc L.
```

```
;  APPLICANT:  Burgess, Catherine E.
;  APPLICANT:  Casman, Stacie J.
;  APPLICANT:  Grosse, William M.
;  APPLICANT:  Gusev, Vladimir Y.
;  APPLICANT:  Ji, Weizhen
;  APPLICANT:  Lepley, Denise M.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Mezick, Amanda J.
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Shenoy, Suresh G.
;  APPLICANT:  Shimkets, Richard A.
;  APPLICANT:  Spaderna, Steven K.
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Szekeres, Edward S. Jr.
;  APPLICANT:  Taupier, Raymond J. Jr.
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Zerhusen, Bryan D.
;  APPLICANT:  Voss, Edward Z.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-050 CIP
;  CURRENT APPLICATION NUMBER: US/10/190,115
;  CURRENT FILING DATE:  2003-02-10
;  PRIOR APPLICATION NUMBER: 60/303,168
;  PRIOR FILING DATE: 2001-07-05
;  PRIOR APPLICATION NUMBER: 60/368,996
;  PRIOR FILING DATE: 2002-04-01
;  PRIOR APPLICATION NUMBER: 60/386,816
;  PRIOR FILING DATE: 2002-06-07
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585,
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2001-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 136
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 29
;   LENGTH: 2531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-190-115-29

  Query Match             1.3%;  Score 9;  DB 15;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 3
US-10-369-072-29
; Sequence 29, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook II, John P
;  APPLICANT:  Spaderna, Stephen K
;  APPLICANT:  Tchernev, Velizar
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Shenoy, Suresh
```

```
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Taupier, Raymond T
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Szerkeres, Edward S
;  APPLICANT:  Lepley, Denise M
;  APPLICANT:  Shen, Lei
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Padigaru, Muralidhara
;  TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-050 CON2
;  CURRENT APPLICATION NUMBER: US/10/369,072
;  CURRENT FILING DATE:  2003-02-18
;  PRIOR APPLICATION NUMBER: 10/174,372
;  PRIOR FILING DATE: 2002-06-17
;  PRIOR APPLICATION NUMBER: 09/898,994
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: 60/215,854
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,856
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/215,902
;  PRIOR FILING DATE: 2000-07-03
;  PRIOR APPLICATION NUMBER: 60/216,585
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,586
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/216,722
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: 60/218,622
;  PRIOR FILING DATE: 2000-07-17
;  PRIOR APPLICATION NUMBER: 60/218,992
;  PRIOR FILING DATE: 2000-07-17
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 100
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
;   LENGTH: 2531
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-369-072-29

  Query Match             1.3%;  Score 9;  DB 15;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 4
US-09-776-724A-136
; Sequence 136, Application US/09776724A
; Publication No. US20030050455A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 64 Human Secreted Proteins
;  FILE REFERENCE: PZ011
;  CURRENT APPLICATION NUMBER: US/09/776,724A
;  CURRENT FILING DATE:  2001-02-06
;  PRIOR APPLICATION NUMBER: 60/180,909
;  PRIOR FILING DATE: 2000-02-08
;  PRIOR APPLICATION NUMBER: 09/669,688
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: 09/229,982
;  PRIOR FILING DATE: 1999-01-14
;  PRIOR APPLICATION NUMBER: PCT/US98/14613
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: 60/052,661
```

```
;  PRIOR FILING DATE: 1997-07-16
;  PRIOR APPLICATION NUMBER: 60/052,872
;  PRIOR FILING DATE: 1997-07-16
;  PRIOR APPLICATION NUMBER: 60/052,871
;  PRIOR FILING DATE: 1997-07-16
;  PRIOR APPLICATION NUMBER: 60/052,874
;  PRIOR FILING DATE: 1997-07-16
;  PRIOR APPLICATION NUMBER: 60/052,873
;  PRIOR FILING DATE: 1997-07-16
;  PRIOR APPLICATION NUMBER: 60/052,870
;  PRIOR FILING DATE: 1997-07-16
;  PRIOR APPLICATION NUMBER: 60/052,875
;  PRIOR FILING DATE: 1997-07-16
;  PRIOR APPLICATION NUMBER: 60/053,440
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/053,441
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/053,442
;  PRIOR FILING DATE: 1997-07-22
;  PRIOR APPLICATION NUMBER: 60/056,359
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,725
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,985
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,952
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,989
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/056,361
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,726
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,724
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,946
;  PRIOR FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: 60/055,683
;  PRIOR FILING DATE: 1997-08-18
;  NUMBER OF SEQ ID NOS: 294
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 136
;   LENGTH: 54
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (32)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (39)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-776-724A-136

  Query Match             1.1%;  Score 8;  DB 10;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         74 VTLLRSLT 81
              ||||||||
Db         47 VTLLRSLT 54

RESULT 5
US-09-864-761-35784
; Sequence 35784, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35784
;   LENGTH: 80
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL035413.19
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =   3
;   OTHER INFORMATION: EST_HUMAN HIT: BE875511.1, EVALUE 3.00e-14
US-09-864-761-35784

  Query Match             1.1%;  Score 8;  DB 9;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        680 PGAPPPQP 687
              ||||||||
Db         33 PGAPPPQP 40

RESULT 6
US-10-264-237-2547
; Sequence 2547, Application US/10264237
```

```
; Publication No. US20040009491A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
;  CURRENT APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2547
;   LENGTH: 131
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (106)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (108)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (115)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (120)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (127)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2547

  Query Match             1.1%;  Score 8;  DB 15;  Length 131;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        278 GVLLFSGG 285
              ||||||||
Db         58 GVLLFSGG 65

RESULT 7
US-09-833-245-2022
; Sequence 2022, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Albumin Fusion Proteins
;  FILE REFERENCE: PF546PCT
;  CURRENT APPLICATION NUMBER: US/09/833,245
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/229, 358
;  PRIOR FILING DATE: 2000-04-12
;  PRIOR APPLICATION NUMBER: 60/256, 931
;  PRIOR FILING DATE: 2000-12-21
;  PRIOR APPLICATION NUMBER: 60/199, 384
;  PRIOR FILING DATE: 2000-04-25
;  NUMBER OF SEQ ID NOS: 2267
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2022
;   LENGTH: 264
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-245-2022

  Query Match             1.1%;  Score 8;  DB 11;  Length 264;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        278 GVLLFSGG 285
              ||||||||
Db         13 GVLLFSGG 20

RESULT 8
US-10-266-829-75
; Sequence 75, Application US/10266829
; Publication No. US20030220489A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 29 Human secreted proteins
;  FILE REFERENCE: PZ041P1
;  CURRENT APPLICATION NUMBER: US/10/266,829
;  CURRENT FILING DATE:  2002-10-09
;  PRIOR APPLICATION NUMBER: 09/756,168
;  PRIOR FILING DATE: 2001-01-09
;  PRIOR APPLICATION NUMBER: PCT/US00/19735
;  PRIOR FILING DATE: 2000-07-20
;  PRIOR APPLICATION NUMBER: 60/145,220
;  PRIOR FILING DATE: 1999-07-23
;  NUMBER OF SEQ ID NOS: 146
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
;   LENGTH: 264
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-266-829-75

  Query Match             1.1%;  Score 8;  DB 15;  Length 264;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        278 GVLLFSGG 285
              ||||||||
Db         13 GVLLFSGG 20

RESULT 9
US-10-219-220-251
; Sequence 251, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
;  APPLICANT: Flinn, Barry
;  APPLICANT:  Lasham, Annette
;  TITLE OF INVENTION: Compositions affecting programmed cell
;  TITLE OF INVENTION:  death and their use in the modification of plant development
;  FILE REFERENCE: 11000.1022c1
;  CURRENT APPLICATION NUMBER: US/10/219,220
;  CURRENT FILING DATE:  2002-08-14
;  PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
;  PRIOR FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 251
;   LENGTH: 299
;   TYPE: PRT
;   ORGANISM: Pinus radiata
US-10-219-220-251

  Query Match             1.1%;  Score 8;  DB 14;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         74 VTLLRSLT 81
              ||||||||
Db        260 VTLLRSLT 267

RESULT 10
```

```
US-10-122-706-31
; Sequence 31, Application US/10122706
; Publication No. US20030119012A1
; GENERAL INFORMATION:
;  APPLICANT: Srinivasan, Maithreyan
;  APPLICANT:  Reifler, Michael
;  TITLE OF INVENTION: Sulfurylase-Luciferase Fusion Proteins
;  FILE REFERENCE: 21465-504
;  CURRENT APPLICATION NUMBER: US/10/122,706
;  CURRENT FILING DATE:  2002-07-01
;  PRIOR APPLICATION NUMBER: 60/335,949
;  PRIOR FILING DATE: 2001-10-30
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
;   LENGTH: 309
;   TYPE: PRT
;   ORGANISM: Thermomonospora fusca
US-10-122-706-31

  Query Match              1.1%;  Score 8;  DB 14;  Length 309;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        279 VLLFSGGK 286
              ||||||||
Db         38 VLLFSGGK 45

RESULT 11
US-09-833-245-770
; Sequence 770, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Albumin Fusion Proteins
;  FILE REFERENCE: PF546PCT
;  CURRENT APPLICATION NUMBER: US/09/833,245
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/229, 358
;  PRIOR FILING DATE: 2000-04-12
;  PRIOR APPLICATION NUMBER: 60/256, 931
;  PRIOR FILING DATE: 2000-12-21
;  PRIOR APPLICATION NUMBER: 60/199, 384
;  PRIOR FILING DATE: 2000-04-25
;  NUMBER OF SEQ ID NOS: 2267
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 770
;   LENGTH: 383
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-245-770

  Query Match              1.1%;  Score 8;  DB 11;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        166 PCQNGGQC 173
              ||||||||
Db        100 PCQNGGQC 107

RESULT 12
US-10-245-103-80
; Sequence 80, Application US/10245103
; Publication No. US20030068778A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin
;  APPLICANT:  Eaton,Dan
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Grimaldi,J. Christopher
;  APPLICANT:  Gurney,Austin
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stephan,Jean-Phillippe
;  APPLICANT:  Watanbe,Colin
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang,Zemin
;  APPLICANT:  Fong,Sherman
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3630R1C112
;  CURRENT APPLICATION NUMBER: US/10/245,103
;  CURRENT FILING DATE:  2002-09-17
;  PRIOR APPLICATION NUMBER: 10/197942
;  PRIOR FILING DATE: 2002-07-18
;  PRIOR APPLICATION NUMBER: 60/059114
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/063046
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/065027
;  PRIOR FILING DATE: 1997-11-10
;  PRIOR APPLICATION NUMBER: 60/079689
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/086478
;  PRIOR FILING DATE: 1998-05-22
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/089801
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/090557
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090689
;  PRIOR FILING DATE: 1998-06-25
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
;   LENGTH: 383
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-245-103-80

  Query Match              1.1%;  Score 8;  DB 14;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        166 PCQNGGQC 173
              ||||||||
Db        100 PCQNGGQC 107

RESULT 13
US-10-245-107-80
; Sequence 80, Application US/10245107
; Publication No. US20030068779A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin
;  APPLICANT:  Eaton,Dan
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Grimaldi,J. Christopher
;  APPLICANT:  Gurney,Austin
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stephan,Jean-Phillippe
;  APPLICANT:  Watanbe,Colin
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang,Zemin
;  APPLICANT:  Fong,Sherman
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3630R1C71
;  CURRENT APPLICATION NUMBER: US/10/245,107
;  CURRENT FILING DATE:  2002-09-16
;  PRIOR APPLICATION NUMBER: 10/197942
;  PRIOR FILING DATE: 2002-07-18
;  PRIOR APPLICATION NUMBER: 60/059114
```

```
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/063046
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/065027
;  PRIOR FILING DATE: 1997-11-10
;  PRIOR APPLICATION NUMBER: 60/079689
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/086478
;  PRIOR FILING DATE: 1998-05-22
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/089801
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/090557
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090689
;  PRIOR FILING DATE: 1998-06-25
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
;   LENGTH: 383
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-245-107-80

  Query Match             1.1%;  Score 8;  DB 14;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        166 PCQNGGQC 173
              ||||||||
Db        100 PCQNGGQC 107


RESULT 14
US-10-245-143-80
; Sequence 80, Application US/10245143
; Publication No. US20030068780A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin
;  APPLICANT:  Eaton,Dan
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Grimaldi,J. Christopher
;  APPLICANT:  Gurney,Austin
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stephan,Jean-Phillippe
;  APPLICANT:  Watanbe,Colin
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang,Zemin
;  APPLICANT:  Fong,Sherman
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3630R1C90
;  CURRENT APPLICATION NUMBER: US/10/245,143
;  CURRENT FILING DATE:  2002-09-16
;  PRIOR APPLICATION NUMBER: 10/197942
;  PRIOR FILING DATE: 2002-07-18
;  PRIOR APPLICATION NUMBER: 60/059114
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/063046
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/065027
;  PRIOR FILING DATE: 1997-11-10
;  PRIOR APPLICATION NUMBER: 60/079689
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/086478
;  PRIOR FILING DATE: 1998-05-22
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/089801
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/090557
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090689
;  PRIOR FILING DATE: 1998-06-25
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
;   LENGTH: 383
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-245-143-80

  Query Match             1.1%;  Score 8;  DB 14;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        166 PCQNGGQC 173
              ||||||||
Db        100 PCQNGGQC 107


RESULT 15
US-10-245-771-80
; Sequence 80, Application US/10245771
; Publication No. US20030068781A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin
;  APPLICANT:  Eaton,Dan
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Grimaldi,J. Christopher
;  APPLICANT:  Gurney,Austin
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stephan,Jean-Phillippe
;  APPLICANT:  Watanbe,Colin
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang,Zemin
;  APPLICANT:  Fong,Sherman
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3630R1C98
;  CURRENT APPLICATION NUMBER: US/10/245,771
;  CURRENT FILING DATE:  2002-09-16
;  PRIOR APPLICATION NUMBER: 10/197942
;  PRIOR FILING DATE: 2002-07-18
;  PRIOR APPLICATION NUMBER: 60/059114
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/063046
;  PRIOR FILING DATE: 1997-10-24
;  PRIOR APPLICATION NUMBER: 60/065027
;  PRIOR FILING DATE: 1997-11-10
;  PRIOR APPLICATION NUMBER: 60/079689
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/086478
;  PRIOR FILING DATE: 1998-05-22
;  PRIOR APPLICATION NUMBER: 60/087607
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/089801
;  PRIOR FILING DATE: 1998-06-18
;  PRIOR APPLICATION NUMBER: 60/090557
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090689
;  PRIOR FILING DATE: 1998-06-25
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
;   LENGTH: 383
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-245-771-80

  Query Match             1.1%;  Score 8;  DB 14;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        166 PCQNGGQC 173
              ||||||||
Db        100 PCQNGGQC 107

Search completed: March  9, 2004, 17:34:05
Job time : 52.7431 secs
```

```
                     GenCore version 5.1.6
             Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:23:39 ; Search time 24.1076 Seconds
                                           (without alignments)
                                           2813.016 Million cell updates/sec

Title:          US-10-006-011A-3
Perfect score:  705
Sequence:       1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       283366 seqs, 96191526 residues

Word size :      0

Total number of hits satisfying chosen parameters:       283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        PIR_78:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                     %
Result              Query
  No.      Score    Match Length DB  ID                       Description
-------------------------------------------------------------------------------
    1       705    100.0   4391  2  A38096                   perlecan precursor
    2        44      6.2   3707  2  S18252                   heparan sulfate pr
    3         9      1.3    861  2  A48825                   Notch homolog Motc
    4         9      1.3   1959  1  AGRT                     agrin - rat
    5         9      1.3   2531  2  S18188                   notch protein homo
    6         9      1.3   2531  2  A46019                   notch-1 protein -
    7         8      1.1    215  2  T48789                   TFS1 related prote
    8         8      1.1    332  2  A70772                   probable sulfate a
    9         8      1.1    430  2  A95935                   conserved hypothet
   10         8      1.1    497  2  AH2015                   sodium/solute symp
   11         8      1.1    500  2  F70012                   leucyl aminopeptid
   12         8      1.1    748  2  T00732                   hypothetical prote
   13         8      1.1    886  2  T39081                   hypothetical prote
   14         8      1.1   1203  2  A49175                   Motch B protein -
   15         8      1.1   1257  2  S28764                   neurocan precursor
   16         8      1.1   1268  2  S52781                   neurocan - mouse
   17         8      1.1   1328  2  T43060                   agrin - electric r
   18         8      1.1   1751  1  MMHUMH                   laminin alpha-2 ch
   19         8      1.1   1955  1  AGCH                     agrin precursor -
   20         8      1.1   2139  2  A35672                   crumbs protein - f
   21         8      1.1   2321  2  S78549                   notch3 protein - h
   22         8      1.1   2471  2  A49128                   cell-fate determin
   23         8      1.1   2531  2  T31070                   notch homolog - se
   24         8      1.1   3106  1  S53868                   laminin alpha-2 ch
   25         7      1.0     46  2  S21922                   T-cell receptor al
   26         7      1.0     93  2  S75571                   hypothetical prote
   27         7      1.0     96  2  E97765                   mutator protein Mu
   28         7      1.0    129  2  T29451                   hypothetical prote
   29         7      1.0    137  2  T43574                   translocation prot
```

```
   30         7      1.0    137  2  B40361                   virC-region hypoth
   31         7      1.0    139  2  S30967                   gene 22 protein -
   32         7      1.0    139  2  B95304                   hypothetical prote
   33         7      1.0    146  2  F75400                   hypothetical prote
   34         7      1.0    154  2  A99404                   hypothetical prote
   35         7      1.0    161  2  B84937                   phosphotransferase
   36         7      1.0    166  2  AB2147                   hypothetical prote
   37         7      1.0    176  2  D71336                   probable peptidyl-
   38         7      1.0    179  2  A33104                   tenebrosin C - sea
   39         7      1.0    181  2  E87418                   chemotaxis protein
   40         7      1.0    196  2  G72718                   probable transcrip
   41         7      1.0    201  2  E75567                   hypothetical prote
   42         7      1.0    210  2  T03144                   hypothetical prote
   43         7      1.0    214  2  JC4682                   equinatoxin II pre
   44         7      1.0    221  2  T24494                   hypothetical prote
   45         7      1.0    224  2  T10660                   photosystem II pro

                                ALIGNMENTS

RESULT 1
A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate protec
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membra
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prot
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515; NID:g29469; PIDN:CAA44373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title; Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989
```

```
A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
anes.
A;Reference number: A33625; MUID:90078352; PMID:2687294
A;Accession: B33625
A;Molecule type: protein
A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A;Accession: A33625
A;Molecule type: protein
A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A;Note: peptide potentially matches four different regions of sequence shown
C;Genetics:
A;Gene: GDB:HSPG2
A;Cross-references: GDB:126372; OMIM:142461
A;Map position: 1p36.1-1p36.1
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-4391/Product: perlecan #status predicted <MAT>
F;22-193/Domain: I <DOM1>
F;194-530/Domain: II <DOM2>
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;531-1676/Domain: III <DOM3>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;1677-3686/Domain: IV <DOM4>
F;2007-2034/Domain: transmembrane #status predicted <TRM>
F;3687-4391/Domain: V <DOM5>
F;3845-3880/Domain: EGF homology <EGF1>
F;3888-3921/Domain: EGF homology <EGF>
F;3953-4106/Domain: laminin G repeat homology <LG2>
F;4147-4175/Domain: EGF homology <EGF2>
F;4149-4151/Region: motor neuron attachment (L-R-E) motif
F;4299-4301/Region: motor neuron attachment (L-R-E) motif
F;65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval
F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

  Query Match             100.0%;  Score 705;  DB 2;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  705;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3687 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3746

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3747 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3806

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3807 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3866

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3867 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3926

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3927 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3987 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4046

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4047 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4106

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4107 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4166

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4167 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4226

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4227 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4286

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4287 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 4346

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4347 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
S18252
heparan sulfate proteoglycan - mouse
N;Alternate names: perlecan
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S18252; A31917; B31917; S66460
R;Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogl
adhesion molecule.
A;Reference number: S18252; MUID:92078153; PMID:1744087
A;Accession: S18252
A;Molecule type: mRNA
A;Residues: 1-3707 <NOO>
A;Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hass
J. Biol. Chem. 263, 16379-16387, 1988
A;Title: Identification of cDNA clones encoding different domains of the basement membra
A;Reference number: A92680; MUID:89034110; PMID:2972708
A;Accession: A31917
A;Molecule type: mRNA
A;Residues: 940-1601 <NO2>
A;Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
A;Accession: B31917
A;Molecule type: mRNA
A;Residues: 1870-2600 <NO3>
A;Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A;Reference number: S66460; MUID:95377282; PMID:7649154
A;Accession: S66460
A;Molecule type: protein
A;Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: glycoprotein
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;764-811/Domain: laminin-type EGF-like homology <LEG>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;3163-3198/Domain: EGF homology <EGF>
F;3270-3423/Domain: laminin G repeat homology <LG2>
F;3464-3492/Domain: EGF homology <EGF7>
F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             6.2%;  Score 44;  DB 2;  Length 3707;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-36;
```

```
  Matches    44;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy         26 SPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRHPTPLALG 69
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3030 SPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRHPTPLALG 3073

RESULT 3
A48825
Notch homolog Motch protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Sep-2002
C;Accession: A48825
R;Reaume, A.G.; Conlon, R.A.; Zirngibl, R.; Yamaguchi, T.P.; Rossant, J.
Dev. Biol. 154, 377-387, 1992
A;Title: Expression analysis of a Notch homologue in the mouse embryo.
A;Reference number: A48825; MUID:93050801; PMID:1426644
A;Accession: A48825
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-861 <REA>
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:119144)
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;26-57/Domain: EGF homology <EGF>
F;64-95/Domain: EGF homology <EGX1>
F;198-229/Domain: EGF homology <EGF2>
F;441-472/Domain: EGF homology <EGX2>

  Query Match             1.3%;  Score 9;  DB 2;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 2.5;
  Matches     9;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db        240 PCLHGGTCQ 248

RESULT 4
AGRT
agrin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0399; A38856
R;Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
Neuron 6, 811-823, 1991
A;Title: Structure and expression of a rat agrin.
A;Reference number: JH0399; MUID:91222570; PMID:1851019
A;Accession: JH0399
A;Molecule type: mRNA
A;Residues: 1-1779;1799-1959 <RUP>
A;Cross-references: GB:M64780; NID:g202798; PIDN:AAA40703.1; PID:g202800
A;Experimental source: embryonic spinal cord
A;Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator
R;Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
J. Neurosci. 12, 3535-3544, 1992
A;Title: Structure and chromosomal localization of the mammalian agrin gene.
A;Reference number: A38856; MUID:92407628; PMID:1326608
A;Accession: A38856
A;Molecule type: mRNA
A;Residues: 1780-1798 <RU2>
A;Cross-references: GB:S44194
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C;Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.
ylcholine receptor clustering activity.
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-1959/Product: agrin, form 1 #status predicted <AG1>
F;1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
F;1-1779,1799-1959/Product: agrin, form 3 #status predicted <AG3>
F;1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
F;1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
F;22-50/Region: hydrophobic
```

```
F;88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
F;163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
F;236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
F;307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
F;381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
F;446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
F;511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
F;540-542/Region: motor neuron attachment (L-R-E) motif
F;596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
F;688-739/Domain: laminin-type EGF-like homology <LE1>
F;742-786/Domain: laminin-type EGF-like homology <LE2>
F;814-864/Domain: Kazal proteinase inhibitor homology <KPI9>
F;869-992/Region: serine/threonine-rich
F;1084-1086/Region: motor neuron attachment (L-R-E) motif
F;1147-1215/Region: serine/threonine-rich
F;1224-1257/Domain: EGF homology <EG1>
F;1287-1442/Domain: laminin G repeat homology <LG1>
F;1444-1476/Domain: EGF homology <EG2>
F;1483-1515/Domain: EGF homology <EG3>
F;1555-1706/Domain: laminin G repeat homology <LG2>
F;1713-1747/Domain: EGF homology <EG4>
F;1807-1959/Domain: laminin G repeat homology <LG3>
F;97-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454-473,
-1476,1483-1494,1488-1504,1506-1515/Disulfide bonds: #status predicted
F;145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             1.3%;  Score 9;  DB 1;  Length 1959;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches     9;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1228 PCLHGGTCQ 1236

RESULT 5
S18188
notch protein homolog - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
C;Accession: S18188
R;Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A;Title: A homolog of Drosophila Notch expressed during mammalian development.
A;Reference number: S18188; MUID:92111383; PMID:1764995
A;Accession: S18188
A;Molecule type: mRNA
A;Residues: 1-2531 <WEI>
A;Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;987-1018/Domain: EGF homology <EGF1>
F;1025-1056/Domain: EGF homology <EGF>
F;1233-1264/Domain: EGF homology <EGF2>
F;1917-1949/Domain: ankyrin repeat homology <AN1>
F;1950-1982/Domain: ankyrin repeat homology <AN2>
F;1984-2016/Domain: ankyrin repeat homology <AN3>
F;2017-2049/Domain: ankyrin repeat homology <AN4>
F;2050-2082/Domain: ankyrin repeat homology <AN5>

  Query Match             1.3%;  Score 9;  DB 2;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches     9;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 6
A46019
notch-1 protein - mouse
N;Alternate names: motch protein
C;Species: Mus musculus (house mouse)
```

```
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Mar-2003
C;Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109
R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid
Genomics 15, 259-264, 1993
A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A;Reference number: A46019; MUID:93194170; PMID:8449489
A;Accession: A46019
A;Status: not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-2531 <DEL>
A;Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503
A;Note: sequence extracted from NCBI backbone (NCBIP:127318)
R;Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
submitted to the EMBL Data Library, April 1992
A;Description: Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggest
A;Reference number: S25144
A;Accession: S25144
A;Molecule type: mRNA
A;Residues: 1551-2108,'Q',2110-2114,'ALP',2118-2170 <FRA>
A;Cross-references: EMBL:Z11886
R;Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety o
A;Reference number: A49175; MUID:93178563; PMID:8440332
A;Accession: C49175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1161-1547 <LAR>
A;Cross-references: EMBL:X68278; NID:g287987; PIDN:CAA48339.1; PID:g287988
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
R;Kopan, R.; Weintraub, H.
J. Cell Biol. 121, 631-641, 1993
A;Title: Mouse notch: expression in hair follicles correlates with cell fate determinati
A;Reference number: A46438; MUID:93252998; PMID:8486742
A;Accession: B46438
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1865-1932,'RR',1935-1937,'L',1938-1967,'I',1969-2044,'IE',2047-2052,'S',2054
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIN:131246, NCBIP:131247)
C;Comment: This protein has many EGF repeats and lin-12[#1172]/Notch repeats.
C;Comment: This protein is one of the neurogenic proteins controlling the decision betwe
C;Genetics:
A;Gene: notch-1
A;Map position: 2
A;Note: proximal region of chromosome 2
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;106-138/Domain: EGF homology <EGF1>
F;144-175/Domain: EGF homology <EG01>
F;222-254/Domain: EGF homology <EGF2>
F;261-292/Domain: EGF homology <EG02>
F;339-370/Domain: EGF homology <EG03>
F;416-449/Domain: EGF homology <EGF3>
F;456-487/Domain: EGF homology <EG04>
F;494-525/Domain: EGF homology <EG05>
F;532-563/Domain: EGF homology <EG06>
F;607-638/Domain: EGF homology <EG07>
F;682-713/Domain: EGF homology <EG08>
F;757-788/Domain: EGF homology <EG09>
F;795-826/Domain: EGF homology <EG10>
F;873-904/Domain: EGF homology <EG11>
F;911-942/Domain: EGF homology <EG12>
F;949-980/Domain: EGF homology <EG13>
F;987-1018/Domain: EGF homology <EG14>
F;1025-1056/Domain: EGF homology <EG15>
F;1063-1094/Domain: EGF homology <EG16>
F;1149-1180/Domain: EGF homology <EG17>
F;1187-1218/Domain: EGF homology <EG18>
F;1233-1264/Domain: EGF homology <EGF4>
F;1352-1383/Domain: EGF homology <EG19>
F;1391-1425/Domain: EGF homology <EGF>
F;1917-1948/Domain: ankyrin repeat homology <AN1>
F;1949-1981/Domain: ankyrin repeat homology <AN2>
F;1983-2015/Domain: ankyrin repeat homology <AN3>
F;2016-2048/Domain: ankyrin repeat homology <AN4>
F;2049-2081/Domain: ankyrin repeat homology <AN5>

  Query Match             1.3%;  Score 9;  DB 2;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 7
T48789
TFS1 related protein [imported] - Neurospora crassa
N;Alternate names: protein 13E11.370
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C;Accession: T48789
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24541
A;Accession: T48789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-215 <SCH>
A;Cross-references: EMBL:AL353820; GSPDB:GN00112; NCSP:13E11.370
A;Experimental source: cosmid contig 13E11; strain 74
C;Genetics:
A;Gene: NCSP:13E11.370
A;Map position: 2

  Query Match             1.1%;  Score 8;  DB 2;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 7.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         90 LAPVNGTS 97
              ||||||||
Db        136 LAPVNGTS 143


RESULT 8
A70772
probable sulfate adenylate transferase subunit 2 - Mycobacterium tuberculosis (strain H3
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70772
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70772
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-332 <COL>
A;Cross-references: GB:Z73419; GB:AL123456; NID:g3261573; PIDN:CAA97751.1; PID:g1322409
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: cysD
C;Superfamily: nodulation protein nodP

  Query Match             1.1%;  Score 8;  DB 2;  Length 332;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        279 VLLFSGGK 286
              ||||||||
Db         64 VLLFSGGK 71
```

```
RESULT 9
A95935
conserved hypothetical membrane protein SMb21241 [imported] - Sinorhizobium meliloti (st
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: A95935
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: A95935
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-430 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC49145.1; PID:g15140630; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMb21241
A;Genome: plasmid

  Query Match             1.1%;  Score 8;  DB 2;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        251 SYLALPAL 258
              ||||||||
Db        208 SYLALPAL 215


RESULT 10
AH2015
sodium/solute symporter [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2015
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2015
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-497 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB78044.1; PID:g17135498; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1678

  Query Match             1.1%;  Score 8;  DB 2;  Length 497;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        551 STASGLLL 558
              ||||||||
Db        325 STASGLLL 332


RESULT 11
F70012
leucyl aminopeptidase homolog yuiE - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Mar-2003
C;Accession: F70012
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F70012
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-500 <KUN>
A;Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15195.1; PID:g2635702
A;Experimental source: strain 168
C;Genetics:
A;Gene: yuiE
C;Superfamily: Cytosol aminopeptidase

  Query Match             1.1%;  Score 8;  DB 2;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        653 DVATLTGG 660
              ||||||||
Db        367 DVATLTGG 374


RESULT 12
T00732
hypothetical protein F22O13.27 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C;Accession: T00732
R;Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con
eologis, A.; Ecker, J.R.
submitted to the EMBL Data Library, April 1998
A;Description: Genomic sequence for Arabidopsis thaliana BAC F22O13.
A;Reference number: Z14200
A;Accession: T00732
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-748 <SHI>
A;Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063465; GSPDB:GN00059; ATSP:F22O1
C;Genetics:
A;Gene: ATSP:F22O13.27
A;Map position: 1
A;Introns: 329/3; 429/3; 460/2

  Query Match             1.1%;  Score 8;  DB 2;  Length 748;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        394 VSVNGKRL 401
              ||||||||
Db        403 VSVNGKRL 410


RESULT 13
T39081
hypothetical protein SPAC7D4.03c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
```

```
C;Accession: T39081
R;Gentles, S.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21826
A;Accession: T39081
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-886 <GEN>
A;Cross-references: EMBL:Z99532; PIDN:CAB16720.1; GSPDB:GN00066; SPDB:SPAC7D4.03c
A;Experimental source: strain 972h-; cosmid c7D4
C;Genetics:
A;Gene: SPDB:SPAC7D4.03c
A;Map position: 1
C;Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC7D4.03c

  Query Match             1.1%;  Score 8;  DB 2;  Length 886;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        394 VSVNGKRL 401
              ||||||||
Db        828 VSVNGKRL 835

RESULT 14
A49175
Motch B protein - mouse (fragment)
N;Alternate names: Notch homolog
C;Species: Mus musculus (house mouse)
C;Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
C;Accession: A49175; PH1570; S32113
R;Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety c
A;Reference number: A49175; MUID:93178563; PMID:8440332
A;Accession: A49175
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1203 <LAR>
A;Cross-references: EMBL:X68279; NID:g287989; PIDN:CAA48340.1; PID:g287990
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126158)
C;Comment: This protein has many EGF repeats and lin-12/Notch repeats.
C;Comment: This protein is one of the neurogenic proteins controlling the decision betwe
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
F;143-174/Domain: EGF homology <EGX1>
F;482-513/Domain: EGF homology <EGF1>
F;560-591/Domain: EGF homology <EGF>
F;674-705/Domain: EGF homology <EGX2>
F;712-743/Domain: EGF homology <EGF3>
F;836-867/Domain: EGF homology <EGX3>

  Query Match             1.1%;  Score 8;  DB 2;  Length 1203;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 PCQHGATC 433
              ||||||||
Db        840 PCQHGATC 847

RESULT 15
S28764
neurocan precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C;Accession: S28764
R;Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A;Title: Cloning and primary structure of neurocan, a developmentally regulated, aggrega
A;Reference number: S28764; MUID:92406907; PMID:1326557
A;Accession: S28764
A;Molecule type: mRNA
```

```
A;Residues: 1-1257 <RAU>
A;Cross-references: EMBL:M97161; NID:g205649; PIDN:AAC37679.1; PID:g205650
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1257/Product: neurocan #status predicted <MAT>
F;176-253/Domain: link protein repeat homology <LNK1>
F;274-355/Domain: link protein repeat homology <LNK2>
F;364-366/Region: cell attachment (R-G-D) motif
F;953-984/Domain: EGF homology <EGF>
F;1029-1149/Domain: C-type lectin homology <LCH>
F;1156-1212/Domain: complement factor H repeat homology <FHD>
F;121,339,737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F;944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

  Query Match             1.1%;  Score 8;  DB 2;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTC 473
              ||||||||
Db        957 PCLHGGTC 964


Search completed: March  9, 2004, 17:31:26
Job time : 26.1076 secs
```

```
                        GenCore version 5.1.6
                Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:21:48 ; Search time 17.0517 Seconds
                                            (without alignments)
                                            2152.832 Million cell updates/sec

Title:           US-10-006-011A-3
Perfect score:   705
Sequence:        1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        141681 seqs, 52070155 residues

Word size :        0

Total number of hits satisfying chosen parameters:       141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result           Query
  No.    Score   Match  Length DB  ID                    Description
-----------------------------------------------------------------------------
    1     705   100.0    4391  1  PGBM_HUMAN            P98160 homo sapien
    2      44     6.2    3707  1  PGBM_MOUSE            Q05793 mus musculu
    3       9     1.3    1959  1  AGRI_RAT              P25304 rattus norv
    4       9     1.3    2531  1  NTC1_MOUSE            Q01705 mus musculu
    5       9     1.3    2531  1  NTC1_RAT              Q07008 rattus norv
    6       8     1.1     309  1  CYSD_MYCTU            Q10599 mycobacteri
    7       8     1.1     500  1  AMPA_BACSU            O32106 bacillus su
    8       8     1.1    1257  1  PGCN_RAT              P55067 rattus norv
    9       8     1.1    1268  1  PGCN_MOUSE            P55066 mus musculu
   10       8     1.1    1321  1  PGCN_HUMAN            O14594 homo sapien
   11       8     1.1    1328  1  AGRI_DISOM            Q90404 discopyge o
   12       8     1.1    1955  1  AGRI_CHICK            P31696 gallus gall
   13       8     1.1    2139  1  CRB_DROME             P10040 drosophila
   14       8     1.1    2321  1  NTC3_HUMAN            Q9um47 homo sapien
   15       8     1.1    2470  1  NTC2_MOUSE            O35516 mus musculu
   16       8     1.1    2471  1  NTC2_HUMAN            Q04721 homo sapien
   17       8     1.1    2471  1  NTC2_RAT              Q9qw30 rattus norv
   18       8     1.1    3106  1  LMA2_MOUSE            Q60675 mus musculu
   19       8     1.1    3110  1  LMA2_HUMAN            P24043 homo sapien
   20       7     1.0      78  1  YO09_BPL2             P42544 bacteriopha
   21       7     1.0     137  1  YSCB_YEREN            Q01243 yersinia en
   22       7     1.0     139  1  VG22_BPML5            Q05228 mycobacteri
   23       7     1.0     161  1  PTGA_BUCAI            Q9wxi7 buchnera ap
   24       7     1.0     176  1  SLYD_TREPA            O83369 treponema p
   25       7     1.0     184  1  RS7_THEAC             Q9hly1 thermoplasm
   26       7     1.0     184  1  RS7_THEVO             Q97cd9 thermoplasm
   27       7     1.0     192  1  RR4B_CYACA            O22029 cyanidium c
   28       7     1.0     214  1  EQT2_ACTEQ            P17723 actinia equ
   29       7     1.0     214  1  EQT4_ACTEQ            Q9y1u9 actinia equ
   30       7     1.0     214  1  EQT5_ACTEQ            Q93109 actinia equ
   31       7     1.0     229  1  MYB1_NEUCR            O13493 neurospora
   32       7     1.0     240  1  XYNC_STRLI            P26220 streptomyce
   33       7     1.0     244  1  YRT2_CAEEL            Q10045 caenorhabdi
   34       7     1.0     248  1  SCTM_HUMAN            Q8wvn6 homo sapien
   35       7     1.0     250  1  VL01_VACCC            P20540 vaccinia vi
   36       7     1.0     250  1  VL01_VACCV            P07612 vaccinia vi
   37       7     1.0     250  1  VL01_VARV             P33040 variola vir
   38       7     1.0     255  1  TFDC_ALCEU            P05403 alcaligenes
   39       7     1.0     273  1  RL2_TREPA             O83222 treponema p
   40       7     1.0     277  1  CHIS_BACSU            O07921 bacillus su
   41       7     1.0     296  1  CYCG_RHOSH            Q53143 rhodobacter
   42       7     1.0     297  1  DHSB_DROME            P21914 drosophila
   43       7     1.0     305  1  FMT_THETH             P43523 thermus the
   44       7     1.0     320  1  MDCF_RHIME            P56949 rhizobium m
   45       7     1.0     325  1  NAS2_ORYSA            Q9feg8 oryza sativ

                              ALIGNMENTS

RESULT 1
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4391 AA.
AC   P98160; Q16287; Q9H3V5;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Skin;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX   MEDLINE=20553141; PubMed=11101850;
RA   Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA   Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA   Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA   Hentati F., Fontaine B.;
RT   "Perlecan, the major proteoglycan of basement membranes, is altered in
RT   patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL   Nat. Genet. 26:480-483(2000).
RN   [4]
RP   SEQUENCE OF 1016-1470 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [5]
RP   SEQUENCE OF 890-1396 FROM N.A.
RC   TISSUE=Fibrosarcoma;
```

```
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [6]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN   [7]
RP   CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX   MEDLINE=22660472; PubMed=12754519;
RA   Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT   "Identification and quantification of N-linked glycoproteins using
RT   hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL   Nat. Biotechnol. 21:660-666(2003).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC       syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC       characterized by permanent myotonia (prolonged failure of muscle
CC       relaxation) and skeletal dysplasia, resulting in reduced stature,
CC       kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; AL445795; CAC18534.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   PIR; A38096; A38096.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
DR   MIM; 142461; -.
DR   MIM; 255800; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
```

```
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00180; EGF_Lam; 12.
DR   SMART; SM00409; IG; 22.
DR   SMART; SM00408; IGc2; 21.
DR   SMART; SM00406; IGv; 7.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 22.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4391       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      198    235       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      284    320       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      324    360       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      367    404       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      405    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1955       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1956   2051       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2052   2151       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2152   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2340       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2341   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2533       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2534   2629       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2630   2726       IG-LIKE C2-TYPE 12.
```

```
  Query Match             100.0%;  Score 705;  DB 1;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  705;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3687 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3746

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3747 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3806

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3807 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3866

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3867 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3926

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3927 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3987 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4046

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4047 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4106

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4107 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4166

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4167 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4226

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4227 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4286

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4287 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 4346

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4347 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
PGBM_MOUSE
ID   PGBM_MOUSE     STANDARD;      PRT;  3707 AA.
AC   Q05793;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=92078153; PubMed=1744087;
RA   Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "The complete sequence of perlecan, a basement membrane heparan
RT   sulfate proteoglycan, reveals extensive similarity with laminin A
RT   chain, low density lipoprotein-receptor, and the neural cell adhesion
RT   molecule.";
RL   J. Biol. Chem. 266:22939-22947(1991).
RN   [2]
RP   SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89034110; PubMed=2972708;
RA   Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "Identification of cDNA clones encoding different domains of the
RT   basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 263:16379-16387(1988).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M77174; AAA39911.1; -.
DR   EMBL; J04054; AAA39899.1; -.
DR   EMBL; J04055; AAA39912.1; -.
DR   PIR; S18252; S18252.
DR   PDB; 1GL4; 28-NOV-01.
DR   MGD; MGI:96257; Hspg2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0008104; P:protein localization; IMP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 15.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 7.
DR   SMART; SM00408; IGc2; 14.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
```

```
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 15.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; 3D-structure.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   3707       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      195    234       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      281    319       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      320    359       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      360    403       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      404    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1954       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1955   2049       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2050   2148       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2149   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2343       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2344   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2532       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2533   2619       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2620   2720       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2721   2809       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2810   2895       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2896   2980       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2984   3162       LAMININ G-LIKE 1.
FT   DOMAIN     3163   3241       EGF-LIKE.
FT   DOMAIN     3245   3425       LAMININ G-LIKE 2.
FT   DOMAIN     3518   3705       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       3615   3617       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    332    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    368    381       BY SIMILARITY.
```

```
FT   DISULFID    375    394       BY SIMILARITY.
FT   DISULFID    388    403       BY SIMILARITY.
FT   DISULFID    428    479       BY SIMILARITY.
FT   DISULFID    764    773       BY SIMILARITY.
FT   DISULFID    766    780       BY SIMILARITY.
FT   DISULFID    783    792       BY SIMILARITY.
FT   DISULFID    795    811       BY SIMILARITY.
FT   DISULFID    814    829       BY SIMILARITY.
FT   DISULFID    816    839       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    869       BY SIMILARITY.
FT   DISULFID   1159   1168       BY SIMILARITY.
FT   DISULFID   1161   1175       BY SIMILARITY.
FT   DISULFID   1178   1187       BY SIMILARITY.
FT   DISULFID   1190   1206       BY SIMILARITY.
FT   DISULFID   1209   1224       BY SIMILARITY.
FT   DISULFID   1211   1234       BY SIMILARITY.
FT   DISULFID   1237   1246       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1275   1287       BY SIMILARITY.
FT   DISULFID   1277   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1307   1322       BY SIMILARITY.
FT   DISULFID   1563   1572       BY SIMILARITY.
FT   DISULFID   1565   1579       BY SIMILARITY.
FT   DISULFID   1582   1591       BY SIMILARITY.
FT   DISULFID   1594   1610       BY SIMILARITY.
FT   DISULFID   1613   1628       BY SIMILARITY.
FT   DISULFID   1615   1638       BY SIMILARITY.
FT   DISULFID   1641   1650       BY SIMILARITY.
FT   DISULFID   1653   1668       BY SIMILARITY.
FT   DISULFID   1792   1839       BY SIMILARITY.
FT   DISULFID   1886   1932       BY SIMILARITY.
FT   DISULFID   1976   2021       BY SIMILARITY.
FT   DISULFID   2073   2118       BY SIMILARITY.
FT   DISULFID   2170   2215       BY SIMILARITY.
FT   DISULFID   2268   2313       BY SIMILARITY.

  Query Match             6.2%;  Score 44;  DB 1;  Length 3707;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-35;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 SPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRHPTPLALG 69
              ||||||||||||||||||||||||||||||||||||||||||||
Db       3030 SPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRHPTPLALG 3073


RESULT 3
AGRI_RAT
ID   AGRI_RAT       STANDARD;      PRT;  1959 AA.
AC   P25304; Q63034;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC   TISSUE=Embryonic spinal cord;
RX   MEDLINE=91222570; PubMed=1851019;
RA   Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
RT   "Structure and expression of a rat agrin.";
RL   Neuron 6:811-823(1991).
RN   [2]
RP   SEQUENCE OF 1777-1801 FROM N.A.
RX   MEDLINE=92407628; PubMed=1326608;
RA   Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
RT   "Structure and chromosomal localization of the mammalian agrin gene.";
RL   J. Neurosci. 12:3535-3544(1992).
```

```
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P25304-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P25304-2; Sequence=VSP_001365;
CC       Name=3;
CC         IsoId=P25304-3; Sequence=VSP_001366;
CC       Name=4;
CC         IsoId=P25304-4; Sequence=VSP_001367;
CC       Name=5;
CC         IsoId=P25304-5; Sequence=VSP_001368;
CC   -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
CC   -!- DEVELOPMENTAL STAGE: More abundant early in development.
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
CC       initiator.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64780; AAA40703.1; -.
DR   EMBL; M64780; AAA40702.1; ALT_INIT.
DR   EMBL; S44194; AAB23326.1; -.
DR   PIR; JH0399; AGRT.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30   1959       AGRIN.
```

```
FT   DOMAIN       65    137       KAZAL-LIKE 1.
FT   DOMAIN      141    212       KAZAL-LIKE 2.
FT   DOMAIN      213    284       KAZAL-LIKE 3.
FT   DOMAIN      287    356       KAZAL-LIKE 4.
FT   DOMAIN      361    429       KAZAL-LIKE 5.
FT   DOMAIN      430    494       KAZAL-LIKE 6.
FT   DOMAIN      495    559       KAZAL-LIKE 7.
FT   DOMAIN      563    645       KAZAL-LIKE 8.
FT   DOMAIN      688    741       LAMININ EGF-LIKE 1.
FT   DOMAIN      742    788       LAMININ EGF-LIKE 2.
FT   DOMAIN      794    864       KAZAL-LIKE 9.
FT   DOMAIN     1023   1145       SEA.
FT   DOMAIN     1220   1258       EGF-LIKE 1.
FT   DOMAIN     1263   1439       LAMININ G-LIKE 1.
FT   DOMAIN     1440   1477       EGF-LIKE 2.
FT   DOMAIN     1479   1516       EGF-LIKE 3.
FT   DOMAIN     1526   1708       LAMININ G-LIKE 2.
FT   DOMAIN     1709   1748       EGF-LIKE 4.
FT   DOMAIN     1784   1956       LAMININ G-LIKE 3.
FT   DOMAIN      869    992       SER/THR-RICH.
FT   DOMAIN     1147   1215       SER/THR-RICH.
FT   DISULFID     97    116       POTENTIAL.
FT   DISULFID    105    137       POTENTIAL.
FT   DISULFID    171    191       POTENTIAL.
FT   DISULFID    180    212       POTENTIAL.
FT   DISULFID    244    263       POTENTIAL.
FT   DISULFID    252    284       POTENTIAL.
FT   DISULFID    316    335       POTENTIAL.
FT   DISULFID    324    356       POTENTIAL.
FT   DISULFID    389    408       POTENTIAL.
FT   DISULFID    397    429       POTENTIAL.
FT   DISULFID    454    473       POTENTIAL.
FT   DISULFID    462    494       POTENTIAL.
FT   DISULFID    518    538       POTENTIAL.
FT   DISULFID    527    559       POTENTIAL.
FT   DISULFID    604    624       POTENTIAL.
FT   DISULFID    613    645       POTENTIAL.
FT   DISULFID    688    700       BY SIMILARITY.
FT   DISULFID    690    707       BY SIMILARITY.
FT   DISULFID    709    718       BY SIMILARITY.
FT   DISULFID    721    739       BY SIMILARITY.
FT   DISULFID    742    754       BY SIMILARITY.
FT   DISULFID    744    761       BY SIMILARITY.
FT   DISULFID    763    772       BY SIMILARITY.
FT   DISULFID    775    786       BY SIMILARITY.
FT   DISULFID    823    843       POTENTIAL.
FT   DISULFID    832    864       POTENTIAL.
FT   DISULFID   1224   1235       BY SIMILARITY.
FT   DISULFID   1229   1246       BY SIMILARITY.
FT   DISULFID   1248   1257       BY SIMILARITY.
FT   DISULFID   1444   1455       POTENTIAL.
FT   DISULFID   1449   1465       POTENTIAL.
FT   DISULFID   1467   1476       POTENTIAL.
FT   DISULFID   1483   1494       BY SIMILARITY.
FT   DISULFID   1488   1504       BY SIMILARITY.
FT   DISULFID   1506   1515       BY SIMILARITY.
FT   DISULFID   1713   1727       BY SIMILARITY.
FT   DISULFID   1721   1736       BY SIMILARITY.
FT   DISULFID   1738   1747       BY SIMILARITY.
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    957    957       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1144   1152       Missing (in isoform 2).
FT                                /FTId=VSP_001365.
FT   VARSPLIC   1780   1798       Missing (in isoform 3).
FT                                /FTId=VSP_001366.
FT   VARSPLIC   1788   1798       Missing (in isoform 4).
FT                                /FTId=VSP_001367.
FT   VARSPLIC   1780   1787       Missing (in isoform 5).
FT                                /FTId=VSP_001368.
FT   VARIANT     314    314       V -> VTCD (IN A VARIANT).
SQ   SEQUENCE   1959 AA;  208645 MW;  7FEFDFDAFF89CC31 CRC64;
```

```
  Query Match             1.3%;  Score 9;  DB 1;  Length 1959;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1228 PCLHGGTCQ 1236


RESULT 4
NTC1_MOUSE
ID   NTC1_MOUSE     STANDARD;      PRT;  2531 AA.
AC   Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9R0X7;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Motch A)
DE   (mT14) (p300).
GN   NOTCH1 OR MOTCH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Embryo;
RX   MEDLINE=93194170; PubMed=8449489;
RA   Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA   Copeland N.G., Gridley T.;
RT   "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT   homolog of Drosophila Notch.";
RL   Genomics 15:259-264(1993).
RN   [2]
RP   SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RC   STRAIN=CD-1; TISSUE=Embryo;
RX   MEDLINE=93050801; PubMed=1426644;
RA   Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
RT   "Expression analysis of a Notch homologue in the mouse embryo.";
RL   Dev. Biol. 154:377-387(1992).
RN   [3]
RP   SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RC   TISSUE=Embryo;
RX   MEDLINE=93048835; PubMed=1425352;
RA   Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA   Greenspan R.J., McMahon A.P., Gridley T.;
RT   "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT   suggests an important role in early postimplantation mouse
RT   development.";
RL   Development 115:737-744(1992).
RN   [4]
RP   SEQUENCE OF 1161-1547 FROM N.A.
RC   STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX   MEDLINE=93178563; PubMed=8440332;
RA   Lardelli M., Lendahl U.;
RT   "Motch A and Motch B-two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   Exp. Cell Res. 204:364-372(1993).
RN   [5]
RP   SEQUENCE OF 1659-1673 FROM N.A.
RX   MEDLINE=99364499; PubMed=10437788;
RA   Lee J.S., Ishimoto A., Yanagawa S.I.;
RT   "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RT   to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
RL   FEBS Lett. 455:276-280(1999).
RN   [6]
RP   SEQUENCE OF 1950-2201 FROM N.A.
RX   MEDLINE=98029496; PubMed=9384671;
RA   Messerle M., Follo M., Nehls M., Eggert H., Boehm T.;
RT   "Dynamic changes in gene expression during in vitro differentiation of
RT   mouse embryonic stem cells.";
RL   Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN   [7]
RP   SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
RP   MUTAGENESIS OF 1651-ARG--ARG-1654.
RX   MEDLINE=98318619; PubMed=9653148;
RA   Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G.,
RA   Israel A.;
RT   "The Notch1 receptor is cleaved constitutively by a furin-like
RT   convertase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
RN   [8]
RP   PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RX   MEDLINE=21523956; PubMed=11518718;
RA   Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT   "Murine notch homologs (N1-4) undergo presenilin-dependent
RT   proteolysis.";
RL   J. Biol. Chem. 276:40268-40273(2001).
RN   [9]
RP   POST-TRANSLATIONAL PROCESSING.
RX   MEDLINE=21374376; PubMed=11459941;
RA   Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT   "Conservation of the biochemical mechanisms of signal transduction
RT   among mammalian Notch family members.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN   [10]
RP   INTERACTION WITH DTX1 AND DTX2.
RX   MEDLINE=21123790; PubMed=11226752;
RA   Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
RA   Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
RA   Okano H., Matsuno K.;
RT   "Murine homologs of deltex define a novel gene family involved in
RT   vertebrate Notch signaling and neurogenesis.";
RL   Int. J. Dev. Neurosci. 19:21-35(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). May play an essential role in
CC       postimplantation development, probably in some aspect of cell
CC       specification and/or differentiation. May be involved in mesoderm
CC       development, somite formation and neurogenesis. Involved in the
CC       maturation of both CD4+ and CD8+ cells in the thymus.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds. Interacts with DTX1 and DTX2.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q01705-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC       thymus. Expressed at lower levels in the spleen, bone-marrow,
CC       spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC       muscle, kidney and heart.
CC   -!- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
CC       8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
CC       endothelial cells, while much lower levels are seen in the
CC       neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
CC       the neuroepithelium. At 13.5 dpc expressed in the surface
CC       ectoderm, eye and developing whisker follicles.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
```

```
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane.
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 36 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z11886; CAA77941.1; -.
DR   EMBL; L02613; AAK14898.1; -.
DR   EMBL; X68278; CAA48339.1; -.
DR   EMBL; AJ238029; CAB40733.1; -.
DR   EMBL; X82562; CAA57909.1; -.
DR   PIR; A46019; A46019.
DR   PIR; B49175; B49175.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:97363; Notch1.
DR   GO; GO:0005887; C:integral to plasma membrane; IC.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0030154; P:cell differentiation; IMP.
DR   GO; GO:0007386; P:compartment specification; IMP.
DR   GO; GO:0007219; P:N signaling pathway; IC.
DR   GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 7.
DR   Pfam; PF00008; EGF; 35.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 24.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 34.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS50026; EGF_3; 36.
DR   PROSITE; PS01187; EGF_CA; 21.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   CHAIN      1711   2531       NOTCH EXTRACELLULAR TRUNCATION.
FT   CHAIN      1744   2531       NOTCH INTRACELLULAR DOMAIN.
FT   DOMAIN       19   1725       EXTRACELLULAR (POTENTIAL).

  Query Match              1.3%;  Score 9;  DB 1;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037

RESULT 5
NTC1_RAT
ID   NTC1_RAT       STANDARD;      PRT;  2531 AA.
AC   Q07008;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN   NOTCH1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Schwann cell;
RX   MEDLINE=92111383; PubMed=1764995;
RA   Weinmaster G., Roberts V.J., Lemke G.;
RT   "A homolog of Drosophila Notch expressed during mammalian
RT   development.";
RL   Development 113:199-205(1991).
RN   [2]
RP   REVISIONS TO 1652-1653.
RA   Weinmaster G.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   FUNCTION.
RX   MEDLINE=21094508; PubMed=11182080;
RA   Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA   Honjo T.;
RT   "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT   neural progenitor cells to an astroglial fate.";
RL   Neuron 29:45-55(2001).
RN   [4]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=93202015; PubMed=1295745;
RA   Weinmaster G., Roberts V.J., Lemke G.;
RT   "Notch2: a second mammalian Notch gene.";
RL   Development 116:931-941(1992).
RN   [5]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=21331789; PubMed=11438922;
RA   Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT   "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT   functional roles for the Notch-DSL signaling system during brain
RT   development.";
RL   J. Comp. Neurol. 436:167-181(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). Acts instructively to control
CC       the cell fate determination of CNS multipotent progenitor cells,
CC       resulting in astroglial induction and neuron/oligodendrocyte
CC       suppression.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
CC       Expressed in postnatal central nervous system (CNS) germinal zones
CC       and, in early postnatal life, within numerous cells throughout the
CC       CNS. Found in both subventricular and ventricular germinal zones.
CC   -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
CC       days 12 and 14 and decrease rapidly to much lower levels in the
CC       adult.
```

```
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 36 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X57405; CAA40667.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 35.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 25.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 26.
DR   PROSITE; PS50026; EGF_3; 36.
DR   PROSITE; PS01187; EGF_CA; 21.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   2531       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT   CHAIN      1711   2531       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1744   2531       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       19   1723       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1724   1746       POTENTIAL.
FT   DOMAIN     1747   2531       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       20     58       EGF-LIKE 1.
FT   DOMAIN       59     99       EGF-LIKE 2.
FT   DOMAIN      102    139       EGF-LIKE 3.
FT   DOMAIN      140    176       EGF-LIKE 4.
FT   DOMAIN      178    216       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      218    255       EGF-LIKE 6.
FT   DOMAIN      257    293       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      295    333       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      335    371       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      372    410       EGF-LIKE 10.
FT   DOMAIN      412    450       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      452    488       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      490    526       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      528    564       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      566    601       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      603    639       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      641    676       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      678    714       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      716    751       EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      753    789       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      791    827       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      829    867       EGF-LIKE 22.
FT   DOMAIN      869    905       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      907    943       EGF-LIKE 24.
FT   DOMAIN      945    981       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      983   1019       EGF-LIKE 26.
FT   DOMAIN     1021   1057       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1059   1095       EGF-LIKE 28.
FT   DOMAIN     1097   1143       EGF-LIKE 29.
FT   DOMAIN     1145   1181       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1183   1219       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1221   1265       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1267   1305       EGF-LIKE 33.
FT   DOMAIN     1307   1346       EGF-LIKE 34.
FT   DOMAIN     1348   1384       EGF-LIKE 35.
FT   DOMAIN     1387   1426       EGF-LIKE 36.
FT   REPEAT     1445   1480       LIN/NOTCH 1.
FT   REPEAT     1481   1522       LIN/NOTCH 2.
FT   REPEAT     1523   1562       LIN/NOTCH 3.
FT   REPEAT     1917   1946       ANK 1.
FT   REPEAT     1950   1980       ANK 2.
FT   REPEAT     1984   2013       ANK 3.
FT   REPEAT     2017   2046       ANK 4.
FT   REPEAT     2050   2079       ANK 5.
FT   DOMAIN     1730   1733       POLY-ALA.
FT   DOMAIN     1891   1894       POLY-GLU.
FT   DOMAIN     2258   2261       POLY-PRO.
FT   DOMAIN     2497   2500       POLY-SER.
FT   SITE       1654   1655       CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY
FT                                SIMILARITY).
FT   DISULFID     24     37       BY SIMILARITY.
FT   DISULFID     31     46       BY SIMILARITY.
FT   DISULFID     48     57       BY SIMILARITY.
FT   DISULFID     63     74       BY SIMILARITY.
FT   DISULFID     68     87       BY SIMILARITY.
FT   DISULFID     89     98       BY SIMILARITY.
FT   DISULFID    106    117       BY SIMILARITY.
FT   DISULFID    111    127       BY SIMILARITY.
FT   DISULFID    129    138       BY SIMILARITY.
FT   DISULFID    144    155       BY SIMILARITY.
FT   DISULFID    149    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    182    195       BY SIMILARITY.
FT   DISULFID    189    204       BY SIMILARITY.
FT   DISULFID    206    215       BY SIMILARITY.
FT   DISULFID    222    233       BY SIMILARITY.
FT   DISULFID    227    243       BY SIMILARITY.

  Query Match             1.3%;  Score 9;  DB 1;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 6
CYSD_MYCTU
ID   CYSD_MYCTU     STANDARD;      PRT;   309 AA.
AC   Q10599;
```

```
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4) (Sulfate adenylate
DE   transferase) (SAT) (ATP-sulfurylase small subunit).
GN   CYSD OR RV1285 OR MT1323 OR MTCY373.04 OR MB1316.
OS   Mycobacterium tuberculosis, and
OS   Mycobacterium bovis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=AF2122/97;
RX   MEDLINE=22709107; PubMed=12788972;
RA   Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA   Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA   Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA   Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT   "The complete genome sequence of Mycobacterium bovis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC   -!- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.
CC   -!- PATHWAY: First step in the sulfate activation pathway. This
CC       reaction occurs early in the reductive branch of the cysteine
CC       biosynthetic pathway.
CC   -!- SUBUNIT: Heterodimer composed of cysD, the smaller subunit, and
CC       cysN (By similarity).
CC   -!- SIMILARITY: Belongs to the PAPS reductase family. CysD subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z73419; CAA97751.1; ALT_INIT.
DR   EMBL; AE007007; AAK45584.1; -.
DR   EMBL; BX248338; CAD94177.1; ALT_INIT.
DR   TIGR; MT1323; -.
DR   TubercuList; Rv1285; -.
DR   HAMAP; MF_00064; -; 1.
DR   InterPro; IPR002500; PAPS_reduct.
DR   Pfam; PF01507; PAPS_reduct; 1.
KW   Cysteine biosynthesis; Transferase; Nucleotidyltransferase;
```

```
KW   Complete proteome.
SQ   SEQUENCE   309 AA;  34888 MW;  75D965D35F1EC284 CRC64;

  Query Match             1.1%;  Score 8;  DB 1;  Length 309;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        279 VLLFSGGK 286
              ||||||||
Db         41 VLLFSGGK 48


RESULT 7
AMPA_BACSU
ID   AMPA_BACSU     STANDARD;      PRT;   500 AA.
AC   O32106;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
DE   (LAP) (Leucyl aminopeptidase).
GN   PEPA OR BSU32050.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: Presumably involved in the processing and regular
CC       turnover of intracellular proteins. Catalyzes the removal of
CC       unsubstituted N-terminal amino acids from various peptides (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
CC       Xbb-, in which Xaa is preferably Leu, but may be other amino acids
CC       including Pro although not Arg or Lys, and Xbb may be Pro.
CC   -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family M17.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z99120; CAB15195.1; -.
DR   PIR; F70012; F70012.
DR   HSSP; P00727; 1LAM.
DR   MEROPS; M17.UPW; -.
DR   SubtiList; BG13970; pepA.
DR   HAMAP; MF_00181; -; 1.
DR   InterPro; IPR000819; Peptidase_M17_C.
DR   InterPro; IPR008283; Peptidase_M17_N.
DR   Pfam; PF00883; Peptidase_M17; 1.
DR   Pfam; PF02789; Peptidase_M17_N; 1.
DR   PRINTS; PR00481; LAMNOPPTDASE.
DR   PROSITE; PS00631; CYTOSOL_AP; 1.
KW   Hydrolase; Aminopeptidase; Manganese; Complete proteome.
FT   METAL       261    261       MANGANESE 2 (BY SIMILARITY).
FT   METAL       266    266       MANGANESE 1 AND 2 (BY SIMILARITY).
FT   METAL       284    284       MANGANESE 2 (BY SIMILARITY).
FT   METAL       343    343       MANGANESE 1 (BY SIMILARITY).
FT   METAL       345    345       MANGANESE 1 AND 2 (BY SIMILARITY).
FT   ACT_SITE    273    273       POTENTIAL.
FT   ACT_SITE    347    347       POTENTIAL.
SQ   SEQUENCE   500 AA;  53657 MW;  3E82968F66566559 CRC64;

  Query Match             1.1%;  Score 8;  DB 1;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        653 DVATLTGG 660
              ||||||||
Db        367 DVATLTGG 374


RESULT 8
PGCN_RAT
ID   PGCN_RAT       STANDARD;      PRT;  1257 AA.
AC   P55067;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3)
DE   (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult
DE   core glycoprotein].
GN   CSPG3 OR NCAN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=92406907; PubMed=1326557;
RA   Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
RT   "Cloning and primary structure of neurocan, a developmentally
RT   regulated, aggregating chondroitin sulfate proteoglycan of brain.";
RL   J. Biol. Chem. 267:19536-19547(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=94230574; PubMed=7513709;
RA   Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
RA   Margolis R.U., Grumet M.;
RT   "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
RT   neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits
RT   neuronal adhesion and neurite outgrowth.";
RL   J. Cell Biol. 125:669-680(1994).
CC   -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC       development by binding to neural cell adhesion molecules (NG-CAM
CC       and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC       acid.
CC   -!- TISSUE SPECIFICITY: Early postnatal and adult brain; not expressed
CC       in kidney, lung, liver and muscle.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC       O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC   -!- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC       DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC       BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M97161; AAC37679.1; -.
DR   PIR; S28764; S28764.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS01241; LINK; 2.
KW   Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW   EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23   1257       NEUROCAN CORE PROTEIN.
FT   CHAIN       639   1257       150 kDa ADULT CORE GLYCOPROTEIN.
FT   DOMAIN       37    157       IG-LIKE V-TYPE.
FT   DOMAIN      158    253       LINK 1.
FT   DOMAIN      259    355       LINK 2.
FT   DOMAIN      949    985       EGF-LIKE 1.
FT   DOMAIN      987   1023       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1025   1154       C-TYPE LECTIN.
FT   DOMAIN     1155   1213       SUSHI.
FT   DISULFID     58    139       BY SIMILARITY.
FT   DISULFID    181    252       BY SIMILARITY.
FT   DISULFID    205    226       BY SIMILARITY.
FT   DISULFID    279    354       BY SIMILARITY.
```

```
FT   DISULFID    303    324       BY SIMILARITY.
FT   DISULFID    953    964       BY SIMILARITY.
FT   DISULFID    958    973       BY SIMILARITY.
FT   DISULFID    975    984       BY SIMILARITY.
FT   DISULFID   1029   1040       BY SIMILARITY.
FT   DISULFID   1057   1149       BY SIMILARITY.
FT   DISULFID   1125   1141       BY SIMILARITY.
FT   DISULFID   1156   1199       BY SIMILARITY.
FT   DISULFID   1185   1212       BY SIMILARITY.
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    339    339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    737    737       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT   CARBOHYD    967    967       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1164   1164       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1257 AA;  135544 MW;  992B33DCFA19EE1B CRC64;

  Query Match             1.1%;  Score 8;  DB 1;  Length 1257;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTC 473
              ||||||||
Db        957 PCLHGGTC 964


RESULT 9
PGCN_MOUSE
ID   PGCN_MOUSE     STANDARD;      PRT;  1268 AA.
AC   P55066;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN   CSPG3 OR NCAN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=96039250; PubMed=7490074;
RA   Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
RA   Faessler R.;
RT   "Structure and chromosomal localization of the mouse neurocan gene.";
RL   Genomics 28:405-410(1995).
CC   -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC       development by binding to neural cell adhesion molecules (NG-CAM
CC       and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC       acid.
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X84727; CAA59216.1; -.
DR   PIR; S52781; S52781.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:104694; Cspg3.
DR   InterPro; IPR002353; AntifreezeII.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
```

```
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS01241; LINK; 2.
KW   Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW   EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   1268       NEUROCAN CORE PROTEIN.
FT   DOMAIN       37    157       IG-LIKE V-TYPE.
FT   DOMAIN      158    253       LINK 1.
FT   DOMAIN      259    355       LINK 2.
FT   DOMAIN      960    996       EGF-LIKE 1.
FT   DOMAIN      998   1034       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1036   1165       C-TYPE LECTIN.
FT   DOMAIN     1166   1224       SUSHI.
FT   DISULFID     58    139       BY SIMILARITY.
FT   DISULFID    181    252       BY SIMILARITY.
FT   DISULFID    205    226       BY SIMILARITY.
FT   DISULFID    279    354       BY SIMILARITY.
FT   DISULFID    303    324       BY SIMILARITY.
FT   DISULFID    964    975       BY SIMILARITY.
FT   DISULFID    969    984       BY SIMILARITY.
FT   DISULFID    986    995       BY SIMILARITY.
FT   DISULFID   1040   1051       BY SIMILARITY.
FT   DISULFID   1068   1160       BY SIMILARITY.
FT   DISULFID   1136   1152       BY SIMILARITY.
FT   DISULFID   1167   1210       BY SIMILARITY.
FT   DISULFID   1196   1223       BY SIMILARITY.
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    339    339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    742    742       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    978    978       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1175   1175       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1268 AA;  137200 MW;  3014E8E202A2FAEC CRC64;

  Query Match             1.1%;  Score 8;  DB 1;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTC 473
              ||||||||
Db        968 PCLHGGTC 975


RESULT 10
PGCN_HUMAN
ID   PGCN_HUMAN     STANDARD;      PRT;  1321 AA.
```

```
AC   O14594; Q9UPK6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN   CSPG3 OR NCAN OR NEUR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99013874; PubMed=9795216;
RA   Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
RT   "Characterization of the human neurocan gene, CSPG3.";
RL   Gene 221:199-205(1998).
RN   [2]
RP   SEQUENCE OF 1-990 AND 1007-1321 FROM N.A., AND VARIANT ALA-1254.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of an ~1 Mb region containing the MEF2B gene in
RT   19p12.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC       development by binding to neural cell adhesion molecules (NG-CAM
CC       and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC       acid.
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF026547; AAC80576.1; -.
DR   EMBL; AC003110; AAB86655.1; -.
DR   EMBL; AC005254; AAC25581.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:2465; CSPG3.
DR   MIM; 600826; -.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   PRINTS; PR01265; LINKMODULE.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS01241; LINK; 2.
KW   Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW   EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal; Polymorphism.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   1321       NEUROCAN CORE PROTEIN.
FT   DOMAIN       38    153       IG-LIKE V-TYPE.
FT   DOMAIN      159    254       LINK 1.
FT   DOMAIN      260    356       LINK 2.
FT   DOMAIN     1008   1044       EGF-LIKE 1.
FT   DOMAIN     1046   1082       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1084   1213       C-TYPE LECTIN.
FT   DOMAIN     1214   1272       SUSHI.
FT   DISULFID     59    140       BY SIMILARITY.
FT   DISULFID    182    253       BY SIMILARITY.
FT   DISULFID    206    227       BY SIMILARITY.
FT   DISULFID    280    355       BY SIMILARITY.
FT   DISULFID    304    325       BY SIMILARITY.
FT   DISULFID   1012   1023       BY SIMILARITY.
FT   DISULFID   1017   1032       BY SIMILARITY.
FT   DISULFID   1034   1043       BY SIMILARITY.
FT   DISULFID   1088   1099       BY SIMILARITY.
FT   DISULFID   1116   1208       BY SIMILARITY.
FT   DISULFID   1184   1200       BY SIMILARITY.
FT   DISULFID   1215   1258       BY SIMILARITY.
FT   DISULFID   1244   1271       BY SIMILARITY.
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1026   1026       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1223   1223       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    1254   1254       V -> A (in dbSNP:1064389).
FT                                /FTId=VAR_016176.
FT   CONFLICT   1234   1234       N -> Y (IN REF. 2).
FT   CONFLICT   1282   1282       G -> R (IN REF. 2).
SQ   SEQUENCE   1321 AA;  142972 MW;  2EF47F823DB980B8 CRC64;

  Query Match             1.1%;  Score 8;  DB 1;  Length 1321;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTC 473
              ||||||||
Db       1016 PCLHGGTC 1023


RESULT 11
AGRI_DISOM
ID   AGRI_DISOM     STANDARD;      PRT;  1328 AA.
AC   Q90404;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin (Fragment).
GN   AGRN.
OS   Discopyge ommata (Electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX   NCBI_TaxID=7785;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
RA   Snow P., McMahan U.J.;
```

```
RT   "Isolation and characterization of a cDNA that encodes an agrin
RT   homolog in the marine ray.";
RL   Mol. Cell. Neurosci. 3:406-417(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction (By similarity).
CC   -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains at least 3 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L01423; AAA49224.1; -.
DR   PIR; T43060; T43060.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 2.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 2.
DR   SMART; SM00280; KAZAL; 2.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT   NON_TER       1      1
FT   DOMAIN       79    132       LAMININ EGF-LIKE 1.
FT   DOMAIN      133    179       LAMININ EGF-LIKE 2.
FT   DOMAIN      411    533       SEA.
FT   DOMAIN      608    644       EGF-LIKE 1.
FT   DOMAIN      649    825       LAMININ G-LIKE 1.
FT   DOMAIN      865    902       EGF-LIKE 2.
FT   DOMAIN      914   1096       LAMININ G-LIKE 2.
FT   DOMAIN     1097   1135       EGF-LIKE 3.
FT   DOMAIN     1146   1324       LAMININ G-LIKE 3.
FT   DISULFID     79     91       BY SIMILARITY.
FT   DISULFID     81     98       BY SIMILARITY.
FT   DISULFID    100    109       BY SIMILARITY.
FT   DISULFID    112    130       BY SIMILARITY.
FT   DISULFID    133    145       BY SIMILARITY.
FT   DISULFID    135    152       BY SIMILARITY.
FT   DISULFID    154    163       BY SIMILARITY.
FT   DISULFID    166    177       BY SIMILARITY.
FT   DISULFID    612    623       BY SIMILARITY.
FT   DISULFID    617    632       BY SIMILARITY.
```

```
FT   DISULFID    634    643       BY SIMILARITY.
FT   DISULFID    869    880       BY SIMILARITY.
FT   DISULFID    874    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID   1101   1114       BY SIMILARITY.
FT   DISULFID   1108   1123       BY SIMILARITY.
FT   DISULFID   1125   1134       BY SIMILARITY.
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1328 AA;  144018 MW;  79D81C1AF2A71C18 CRC64;

  Query Match             1.1%;  Score 8;  DB 1;  Length 1328;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTC 473
              ||||||||
Db        616 PCLHGGTC 623


RESULT 12
AGRI_CHICK
ID   AGRI_CHICK     STANDARD;      PRT;  1955 AA.
AC   P31696;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92232297; PubMed=1314620;
RA   Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahan U.J.;
RT   "cDNA that encodes active agrin.";
RL   Neuron 8:677-689(1992).
RN   [2]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=92232298; PubMed=1314621;
RA   Ruegg M.A., Tsim K.W.K., Horton S.E., Kroeger S., Escher G.,
RA   Gensch E.M., McMahan U.J.;
RT   "The agrin gene codes for a family of basal lamina proteins that
RT   differ in function and distribution.";
RL   Neuron 8:691-699(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P31696-1; Sequence=Displayed;
CC       Name=2; Synonyms=Agrin-related protein 1;
CC         IsoId=P31696-2; Sequence=VSP_001370;
CC       Name=3; Synonyms=Agrin-related protein 2;
CC         IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
```

```
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M94271; AAA48585.1; -.
DR   EMBL; M97371; AAA48586.1; -.
DR   EMBL; M97372; -; NOT_ANNOTATED_CDS.
DR   PIR; JH0591; AGCH.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1955       AGRIN.
FT   DOMAIN       54    126       KAZAL-LIKE 1.
FT   DOMAIN      130    201       KAZAL-LIKE 2.
FT   DOMAIN      202    273       KAZAL-LIKE 3.
FT   DOMAIN      276    344       KAZAL-LIKE 4.
FT   DOMAIN      350    418       KAZAL-LIKE 5.
FT   DOMAIN      419    483       KAZAL-LIKE 6.
FT   DOMAIN      484    548       KAZAL-LIKE 7.
FT   DOMAIN      551    633       KAZAL-LIKE 8.
FT   DOMAIN      675    728       LAMININ EGF-LIKE 1.
FT   DOMAIN      729    775       LAMININ EGF-LIKE 2.
FT   DOMAIN      781    851       KAZAL-LIKE 9.
FT   DOMAIN     1026   1148       SEA.
FT   DOMAIN     1229   1265       EGF-LIKE 1.
FT   DOMAIN     1270   1445       LAMININ G-LIKE 1.
FT   DOMAIN     1446   1483       EGF-LIKE 2.
FT   DOMAIN     1485   1522       EGF-LIKE 3.
FT   DOMAIN     1532   1713       LAMININ G-LIKE 2.
FT   DOMAIN     1714   1752       EGF-LIKE 4.
FT   DOMAIN     1776   1952       LAMININ G-LIKE 3.
FT   DOMAIN      856    995       SER/THR-RICH.
FT   DOMAIN     1150   1219       SER/THR-RICH.
FT   DISULFID     86    105       POTENTIAL.
FT   DISULFID     94    126       POTENTIAL.
```

```
FT   DISULFID    160    180       POTENTIAL.
FT   DISULFID    169    201       POTENTIAL.
FT   DISULFID    233    252       POTENTIAL.
FT   DISULFID    241    273       POTENTIAL.
FT   DISULFID    304    323       POTENTIAL.
FT   DISULFID    312    344       POTENTIAL.
FT   DISULFID    378    397       POTENTIAL.
FT   DISULFID    386    418       POTENTIAL.
FT   DISULFID    443    462       POTENTIAL.
FT   DISULFID    451    483       POTENTIAL.
FT   DISULFID    507    527       POTENTIAL.
FT   DISULFID    516    548       POTENTIAL.
FT   DISULFID    592    612       POTENTIAL.
FT   DISULFID    601    633       POTENTIAL.
FT   DISULFID    675    687       BY SIMILARITY.
FT   DISULFID    677    694       BY SIMILARITY.
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    708    726       BY SIMILARITY.
FT   DISULFID    729    741       BY SIMILARITY.
FT   DISULFID    731    748       BY SIMILARITY.
FT   DISULFID    750    759       BY SIMILARITY.
FT   DISULFID    762    773       BY SIMILARITY.
FT   DISULFID    810    830       POTENTIAL.
FT   DISULFID    819    851       POTENTIAL.
FT   DISULFID   1233   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1450   1461       BY SIMILARITY.
FT   DISULFID   1455   1471       BY SIMILARITY.
FT   DISULFID   1473   1482       BY SIMILARITY.
FT   DISULFID   1489   1500       BY SIMILARITY.
FT   DISULFID   1494   1510       BY SIMILARITY.
FT   DISULFID   1512   1521       BY SIMILARITY.
FT   DISULFID   1718   1731       BY SIMILARITY.
FT   DISULFID   1725   1740       BY SIMILARITY.
FT   DISULFID   1742   1751       BY SIMILARITY.
FT   CARBOHYD    390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1648   1651       Missing (in isoform 3).
FT                                /FTId=VSP_001369.
FT   VARSPLIC   1783   1793       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_001370.
FT   CONFLICT   1129   1131       RTI -> SIL (IN REF. 1; AAA48586).
SQ   SEQUENCE   1955 AA;  211411 MW;  B4DEB27C23422581 CRC64;

  Query Match             1.1%;  Score 8;  DB 1;  Length 1955;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTC 473
              ||||||||
Db       1237 PCLHGGTC 1244


RESULT 13
CRB_DROME
ID   CRB_DROME      STANDARD;      PRT;  2139 AA.
AC   P10040;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Crumbs protein precursor (95F).
GN   CRB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R; TISSUE=Embryo;
```

```
RX   MEDLINE=90263104; PubMed=2344615;
RA   Tepass U., Theres C., Knust E.;
RT   "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT   Drosophila epithelial cells and required for organization of
RT   epithelia.";
RL   Cell 61:787-799(1990).
RN   [2]
RP   SEQUENCE OF 1663-1955 FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=87218537; PubMed=3107986;
RA   Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,
RA   Vaessin H., Campos-Ortega J.A.;
RT   "EGF homologous sequences encoded in the genome of Drosophila
RT   melanogaster, and their relation to neurogenic genes.";
RL   EMBO J. 6:761-766(1987).
CC   -!- FUNCTION: May play a role in the development of epithelia,
CC       possibly for the establishment and/or maintenance of cell
CC       polarity. It may act as a signal.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC   -!- SIMILARITY: Contains 29 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M33753; AAA28428.1; ALT_SEQ.
DR   EMBL; X05144; CAA28793.1; -.
DR   PIR; A35672; A35672.
DR   PIR; B26637; B26637.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0000368; crb.
DR   GO; GO:0016324; C:apical plasma membrane; NAS.
DR   GO; GO:0016327; C:apicolateral plasma membrane; IDA.
DR   GO; GO:0007163; P:establishment and/or maintenance of cell po. . .; IMP.
DR   GO; GO:0016332; P:establishment and/or maintenance of polarit. . .; IMP.
DR   GO; GO:0016334; P:establishment and/or maintenance of polarit. . .; IMP.
DR   GO; GO:0045494; P:photoreceptor maintenance; IMP.
DR   GO; GO:0042052; P:rhabdomere development; NAS.
DR   GO; GO:0045186; P:zonula adherens assembly; IMP.
DR   GO; GO:0045218; P:zonula adherens maintenance; IMP.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 26.
DR   Pfam; PF00054; laminin_G; 3.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00179; EGF_CA; 11.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 15.
DR   PROSITE; PS00022; EGF_1; 26.
DR   PROSITE; PS01186; EGF_2; 17.
DR   PROSITE; PS50026; EGF_3; 27.
DR   PROSITE; PS01187; EGF_CA; 12.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   Differentiation; Repeat; EGF-like domain; Transmembrane;
KW   Glycoprotein; Signal; Phosphorylation.
FT   SIGNAL        1     90
FT   CHAIN        91   2139       CRUMBS PROTEIN.
FT   DOMAIN       91   2084       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2085   2111       POTENTIAL.
FT   DOMAIN     2112   2139       CYTOPLASMIC (POTENTIAL).
```

```
FT   DOMAIN      267    303       EGF-LIKE 1.
FT   DOMAIN      306    343       EGF-LIKE 2.
FT   DOMAIN      348    386       EGF-LIKE 3.
FT   DOMAIN      388    425       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      427    463       EGF-LIKE 5.
FT   DOMAIN      464    500       EGF-LIKE 6.
FT   DOMAIN      501    532       EGF-LIKE 7.
FT   DOMAIN      545    581       EGF-LIKE 8.
FT   DOMAIN      582    611       EGF-LIKE 9.
FT   DOMAIN      609    646       EGF-LIKE 10.
FT   DOMAIN      648    685       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      687    723       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      725    761       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      763    800       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      802    838       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      840    902       EGF-LIKE 16.
FT   DOMAIN      904    940       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      942    978       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      980   1021       EGF-LIKE 19.
FT   DOMAIN     1023   1205       LAMININ G-LIKE 1.
FT   DOMAIN     1207   1243       EGF-LIKE 20.
FT   DOMAIN     1250   1480       LAMININ G-LIKE 2.
FT   DOMAIN     1481   1517       EGF-LIKE 21.
FT   DOMAIN     1558   1758       LAMININ G-LIKE 3.
FT   DOMAIN     1759   1795       EGF-LIKE 22.
FT   DOMAIN     1797   1833       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1835   1871       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1874   1915       EGF-LIKE 25.
FT   DOMAIN     1915   1951       EGF-LIKE 26.
FT   DOMAIN     1953   1989       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1991   2029       EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     2030   2070       EGF-LIKE 29.
FT   DISULFID    271    282       BY SIMILARITY.
FT   DISULFID    276    291       BY SIMILARITY.
FT   DISULFID    293    302       BY SIMILARITY.
FT   DISULFID    310    321       BY SIMILARITY.
FT   DISULFID    315    331       BY SIMILARITY.
FT   DISULFID    333    342       BY SIMILARITY.
FT   DISULFID    352    363       BY SIMILARITY.
FT   DISULFID    357    374       BY SIMILARITY.
FT   DISULFID    376    385       BY SIMILARITY.
FT   DISULFID    392    403       BY SIMILARITY.
FT   DISULFID    397    412       BY SIMILARITY.
FT   DISULFID    414    424       BY SIMILARITY.
FT   DISULFID    431    442       BY SIMILARITY.
FT   DISULFID    436    451       BY SIMILARITY.
FT   DISULFID    453    462       BY SIMILARITY.
FT   DISULFID    468    479       BY SIMILARITY.
FT   DISULFID    473    488       BY SIMILARITY.
FT   DISULFID    490    499       BY SIMILARITY.
FT   DISULFID    505    515       BY SIMILARITY.
FT   DISULFID    509    520       BY SIMILARITY.
FT   DISULFID    522    531       BY SIMILARITY.
FT   DISULFID    549    562       BY SIMILARITY.
FT   DISULFID    556    569       BY SIMILARITY.
FT   DISULFID    571    580       BY SIMILARITY.
FT   DISULFID    586    597       BY SIMILARITY.
FT   DISULFID    591    602       BY SIMILARITY.
FT   DISULFID    604    610       BY SIMILARITY.
FT   DISULFID    613    624       BY SIMILARITY.
FT   DISULFID    618    634       BY SIMILARITY.
FT   DISULFID    636    645       BY SIMILARITY.
FT   DISULFID    652    664       BY SIMILARITY.
FT   DISULFID    659    673       BY SIMILARITY.
FT   DISULFID    675    684       BY SIMILARITY.
FT   DISULFID    691    702       BY SIMILARITY.
FT   DISULFID    696    711       BY SIMILARITY.
FT   DISULFID    713    722       BY SIMILARITY.
FT   DISULFID    729    740       BY SIMILARITY.
FT   DISULFID    734    749       BY SIMILARITY.
FT   DISULFID    751    760       BY SIMILARITY.
FT   DISULFID    767    778       BY SIMILARITY.
FT   DISULFID    772    787       BY SIMILARITY.
```

```
FT   DISULFID    789    799       BY SIMILARITY.
FT   DISULFID    806    817       BY SIMILARITY.
FT   DISULFID    811    826       BY SIMILARITY.
FT   DISULFID    828    837       BY SIMILARITY.
FT   DISULFID    844    855       BY SIMILARITY.
FT   DISULFID    849    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID    908    919       BY SIMILARITY.
FT   DISULFID    913    928       BY SIMILARITY.
FT   DISULFID    930    939       BY SIMILARITY.
FT   DISULFID    946    957       BY SIMILARITY.
FT   DISULFID    952    966       BY SIMILARITY.
FT   DISULFID    968    977       BY SIMILARITY.
FT   DISULFID    984    995       BY SIMILARITY.
FT   DISULFID    989   1009       BY SIMILARITY.
FT   DISULFID   1011   1020       BY SIMILARITY.
FT   DISULFID   1211   1222       BY SIMILARITY.
FT   DISULFID   1216   1231       BY SIMILARITY.
FT   DISULFID   1233   1242       BY SIMILARITY.
FT   DISULFID   1485   1496       BY SIMILARITY.
FT   DISULFID   1490   1505       BY SIMILARITY.
FT   DISULFID   1507   1516       BY SIMILARITY.
FT   DISULFID   1763   1774       BY SIMILARITY.
FT   DISULFID   1768   1783       BY SIMILARITY.
FT   DISULFID   1785   1794       BY SIMILARITY.
FT   DISULFID   1801   1812       BY SIMILARITY.
FT   DISULFID   1806   1821       BY SIMILARITY.
FT   DISULFID   1823   1832       BY SIMILARITY.
FT   DISULFID   1839   1850       BY SIMILARITY.
FT   DISULFID   1844   1859       BY SIMILARITY.
FT   DISULFID   1861   1870       BY SIMILARITY.
FT   DISULFID   1878   1889       BY SIMILARITY.
FT   DISULFID   1883   1903       BY SIMILARITY.
FT   DISULFID   1905   1914       BY SIMILARITY.
FT   DISULFID   1919   1930       BY SIMILARITY.
FT   DISULFID   1924   1939       BY SIMILARITY.
FT   DISULFID   1941   1950       BY SIMILARITY.
FT   DISULFID   1957   1968       BY SIMILARITY.
FT   DISULFID   1962   1977       BY SIMILARITY.
  Query Match             1.1%;  Score 8;  DB 1;  Length 2139;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        166 PCQNGGQC 173
              ||||||||
Db       1961 PCQNGGQC 1968


RESULT 14
NTC3_HUMAN
ID   NTC3_HUMAN     STANDARD;      PRT;  2321 AA.
AC   Q9UM47; Q9UEB3; Q9UPL3; Q9Y6L8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN   NOTCH3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97032728; PubMed=8878478;
RA   Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
RA   Alamowitch S., Domenga V., Cecillion M., Marechal E., Maciazek J.,
RA   Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
RA   Bach J.-F., Bousser M.-G., Tournier-Lasserve E.;
RT   "Notch3 mutations in CADASIL, a hereditary adult-onset condition
RT   causing stroke and dementia.";
RL   Nature 383:707-710(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Gunel M., Artavanis-Tsakonas S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,
RA   Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA   Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA   Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
RA   Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA   Carrano A.V.;
RT   "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
RT   19p13.1.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
RP   ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
RP   GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
RP   CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
RP   ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
RX   MEDLINE=98049753; PubMed=9388399;
RA   Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
RA   Vayssiere C., Cruaud C., Maciazek J., Weissenbach J., Bousser M.-G.,
RA   Bach J.-F., Tournier-Lasserve E.;
RT   "Strong clustering and stereotyped nature of Notch3 mutations in
RT   CADASIL patients.";
RL   Lancet 350:1511-1515(1997).
RN   [5]
RP   VARIANT CADASIL 114-GLY--PRO-120 DEL.
RX   MEDLINE=20264473; PubMed=10802807;
RA   Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
RA   Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserve E.;
RT   "Splice site mutation causing a seven amino acid Notch3 in-frame
RT   deletion in CADASIL.";
RL   Neurology 54:1874-1875(2000).
RN   [6]
RP   IDENTIFICATION OF LIGANDS.
RX   MEDLINE=99180765; PubMed=10079256;
RA   Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA   Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT   "Human ligands of the Notch receptor.";
RL   Am. J. Pathol. 154:785-794(1999).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity).
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
CC       tissues.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
CC       dominant arteriopathy with subcortical infarcts and
CC       leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type
```

```
CC       of stroke and dementia of which key features include recurrent
CC       subcortical ischemic events and vascular dementia.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 34 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U97669; AAB91371.1; -.
DR   EMBL; AF058900; AAC14346.1; -.
DR   EMBL; AF058881; AAC14346.1; JOINED.
DR   EMBL; AF058882; AAC14346.1; JOINED.
DR   EMBL; AF058883; AAC14346.1; JOINED.
DR   EMBL; AF058884; AAC14346.1; JOINED.
DR   EMBL; AF058885; AAC14346.1; JOINED.
DR   EMBL; AF058886; AAC14346.1; JOINED.
DR   EMBL; AF058887; AAC14346.1; JOINED.
DR   EMBL; AF058888; AAC14346.1; JOINED.
DR   EMBL; AF058889; AAC14346.1; JOINED.
DR   EMBL; AF058890; AAC14346.1; JOINED.
DR   EMBL; AF058891; AAC14346.1; JOINED.
DR   EMBL; AF058892; AAC14346.1; JOINED.
DR   EMBL; AF058893; AAC14346.1; JOINED.
DR   EMBL; AF058894; AAC14346.1; JOINED.
DR   EMBL; AF058895; AAC14346.1; JOINED.
DR   EMBL; AF058896; AAC14346.1; JOINED.
DR   EMBL; AF058897; AAC14346.1; JOINED.
DR   EMBL; AF058898; AAC14346.1; JOINED.
DR   EMBL; AF058899; AAC14346.1; JOINED.
DR   EMBL; AC004257; AAC04897.1; -.
DR   EMBL; AC004663; AAC15789.1; ALT_INIT.
DR   PIR; S78549; S78549.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:7883; NOTCH3.
DR   MIM; 600276; -.
DR   MIM; 125310; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 34.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 19.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 25.
DR   PROSITE; PS50026; EGF_3; 34.
DR   PROSITE; PS01187; EGF_CA; 16.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW   Disease mutation.
FT   SIGNAL        1     39       POTENTIAL.
FT   CHAIN        40   2321       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT   CHAIN      1629   2321       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1662   2321       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       40   1643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1644   1664       POTENTIAL.
FT   DOMAIN     1665   2321       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       40     77       EGF-LIKE 1.
FT   DOMAIN       78    118       EGF-LIKE 2.
FT   DOMAIN      119    156       EGF-LIKE 3.
FT   DOMAIN      158    195       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      197    234       EGF-LIKE 5.
FT   DOMAIN      236    272       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      274    312       EGF-LIKE 7.
FT   DOMAIN      314    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      351    389       EGF-LIKE 9.
FT   DOMAIN      391    429       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      431    467       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      469    505       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      507    543       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      545    580       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      582    618       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      620    655       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      657    693       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      695    730       EGF-LIKE 18.
FT   DOMAIN      734    770       EGF-LIKE 19.
FT   DOMAIN      771    808       EGF-LIKE 20.
FT   DOMAIN      810    847       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      849    885       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      887    922       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).

  Query Match             1.1%;  Score 8;  DB 1;  Length 2321;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        483 GFSGPRCQ 490
              ||||||||
Db       1237 GFSGPRCQ 1244


RESULT 15
NTC2_MOUSE
ID   NTC2_MOUSE     STANDARD;      PRT;  2470 AA.
AC   O35516; Q06008; Q60941;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Motch
DE   B).
GN   NOTCH2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Thymus;
RA   Hamada Y., Higuchi M., Tsujimoto Y.;
RT   "Complete amino acid sequence and mutliform transcripts encoded by a
RT   single copy of mouse Notch2 gene.";
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 316-1518 FROM N.A.
RC   STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX   MEDLINE=93178563; PubMed=8440332;
RA   Lardelli M., Lendahl U.;
RT   "Motch A and Motch B-two mouse Notch homologues coexpressed in a
RT   wide variety of tissues.";
RL   Exp. Cell Res. 204:364-372(1993).
RN   [3]
RP   SEQUENCE OF 1765-2153 FROM N.A.
```

```
RX   MEDLINE=97075110; PubMed=8917536;
RA   Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA   Martin D.I.;
RT   "Inhibition of granulocytic differentiation by mNotch1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=99396706; PubMed=10393120;
RA   Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA   Tsujimoto Y.;
RT   "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT   embryonic lethality.";
RL   Development 126:3415-3424(1999).
RN   [5]
RP   DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX   MEDLINE=95333893; PubMed=7609614;
RA   Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT   "Differential expression of Notch1 and Notch2 in developing and adult
RT   mouse brain.";
RL   Brain Res. Mol. Brain Res. 29:263-272(1995).
RN   [6]
RP   POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX   MEDLINE=21523956; PubMed=11518718;
RA   Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT   "Murine notch homologs (N1-4) undergo presenilin-dependent
RT   proteolysis.";
RL   J. Biol. Chem. 276:40268-40273(2001).
RN   [7]
RP   POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX   MEDLINE=21374376; PubMed=11459941;
RA   Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT   "Conservation of the biochemical mechanisms of signal transduction
RT   among mammalian Notch family members.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). May play an essential role in
CC       postimplantation development, probably in some aspect of cell
CC       specification and/or differentiation.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=O35516-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O35516-2; Sequence=VSP_001405;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
CC       neuroepithelia, somites, optic vesicles and branchial arches, but
CC       not heart.
CC   -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
CC       the postnatal ependymal cells, and the choroid plexus throughout
CC       embryonic and postnatal development.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane.
CC   -!- PTM: Phosphorylated.
```

```
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 35 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 6 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D32210; BAA22094.1; -.
DR   EMBL; X68279; CAA48340.1; -.
DR   EMBL; U31881; AAC52924.1; -.
DR   PIR; A49175; A49175.
DR   HSSP; P16109; 1FSB.
DR   MGD; MGI:97364; Notch2.
DR   GO; GO:0005887; C:integral to plasma membrane; IC.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0002011; P:morphogenesis of an epithelial sheet; IMP.
DR   GO; GO:0007219; P:N signaling pathway; IC.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 34.
DR   Pfam; PF00066; notch; 2.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 23.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS50026; EGF_3; 35.
DR   PROSITE; PS01187; EGF_CA; 22.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW   Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   2470       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
FT   CHAIN      1666   2470       NOTCH EXTRACELLULAR TRUNCATION.
FT   CHAIN      1697   2470       NOTCH INTRACELLULAR DOMAIN.
FT   DOMAIN       26   1677       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1678   1698       POTENTIAL.
FT   DOMAIN     1699   2470       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       26     63       EGF-LIKE 1.
FT   DOMAIN       64    102       EGF-LIKE 2.
FT   DOMAIN      105    143       EGF-LIKE 3.
FT   DOMAIN      144    180       EGF-LIKE 4.
FT   DOMAIN      182    219       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      221    256       EGF-LIKE 6 (INCOMPLETE).
FT   DOMAIN      258    294       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      296    334       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      336    372       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      373    411       EGF-LIKE 10.
FT   DOMAIN      413    452       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      454    490       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      492    528       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      530    566       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      568    603       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      605    641       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      643    678       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      680    716       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      718    753       EGF-LIKE 19.
FT   DOMAIN      755    791       EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      793    829       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      831    869       EGF-LIKE 22.
FT   DOMAIN      871    907       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      909    945       EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      947    983       EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      985   1021       EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1023   1059       EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1061   1097       EGF-LIKE 28.
FT   DOMAIN     1099   1145       EGF-LIKE 29.
FT   DOMAIN     1147   1183       EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1185   1221       EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1223   1260       EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1262   1300       EGF-LIKE 33.
FT   DOMAIN     1302   1345       EGF-LIKE 34.
FT   DOMAIN     1372   1410       EGF-LIKE 35.
FT   REPEAT     1418   1454       LIN/NOTCH 1.
FT   REPEAT     1501   1533       LIN/NOTCH 2.
FT   REPEAT     1825   1869       ANK 1.

  Query Match              1.1%;  Score 8;  DB 1;  Length 2470;
  Best Local Similarity   100.0%;  Pred. No. 50;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        426 PCQHGATC 433
              ||||||||
Db       1155 PCQHGATC 1162

Search completed: March  9, 2004, 17:28:34
Job time : 19.0517 secs
```

```
                         GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          March  9, 2004, 17:22:43 ; Search time 61.7389 Seconds
                                           (without alignments)
                                           3602.917 Million cell updates/sec

Title:           US-10-006-011A-3
Perfect score:   705
Sequence:        1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:   OLIGO
                 Gapop 60.0 , Gapext 60.0

Searched:        1017041 seqs, 315518202 residues

Word size :       0

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :        SPTREMBL_25:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                       Description
-----------------------------------------------------------------------
    1     13    1.8      68  11  O08591                  O08591 rattus norv
    2     10    1.4      53  13  Q9DFL0                  Q9dfl0 gillichthys
    3      9    1.3    1208  11  Q80YA8                  Q80ya8 mus musculu
    4      9    1.3    2516  11  Q7TQ52                  Q7tq52 mus musculu
    5      9    1.3    2526  11  Q7TQ51                  Q7tq51 mus musculu
    6      9    1.3    2531  11  Q8K428                  Q8k428 mus musculu
    7      9    1.3    2531  11  Q7TQ50                  Q7tq50 mus musculu
    8      9    1.3    3367  5   Q9XZC9                 Q9xzc9 drosophila
    9      9    1.3    3375  5   Q8IP51                 Q8ip51 drosophila
   10      8    1.1      97  5   Q8I718                 Q8i718 plasmodium
   11      8    1.1     219  11  Q8BMI5                  Q8bmi5 mus musculu
   12      8    1.1     244  3   Q9P6X9                 Q9p6x9 neurospora
   13      8    1.1     302  16  Q82W41                  Q82w41 nitrosomona
   14      8    1.1     308  2   Q842M1                 Q842m1 rhodococcus
   15      8    1.1     320  16  Q8G4F5                  Q8g4f5 bifidobacte
   16      8    1.1     382  11  Q8K1E3                  Q8k1e3 mus musculu
```

```
   17       8    1.1    388  5   Q7YWC2                  Q7ywc2 toxoplasma
   18       8    1.1    430  16  Q92VG0                   Q92vg0 rhizobium m
   19       8    1.1    487  5   Q8MSX5                  Q8msx5 drosophila
   20       8    1.1    487  16  Q9FBM8                   Q9fbm8 streptomyce
   21       8    1.1    489  4   Q8N1J0                  Q8n1j0 homo sapien
   22       8    1.1    494  16  Q81XS5                   Q81xs5 bacillus an
   23       8    1.1    494  16  Q816E3                   Q816e3 bacillus ce
   24       8    1.1    497  2   Q7X2C8                  Q7x2c8 geobacillus
   25       8    1.1    497  16  Q8YWD5                   Q8ywd5 anabaena sp
   26       8    1.1    615  13  O57409                   O57409 brachydanio
   27       8    1.1    642  13  P79941                   P79941 xenopus lae
   28       8    1.1    642  13  Q7ZXT4                   Q7zxt4 xenopus lae
   29       8    1.1    659  11  Q7TQI9                   Q7tqi9 mus musculu
   30       8    1.1    662  4   Q9H8X6                  Q9h8x6 homo sapien
   31       8    1.1    664  13  Q9IAT6                   Q9iat6 brachydanio
   32       8    1.1    702  4   Q9HA02                  Q9ha02 homo sapien
   33       8    1.1    731  4   Q9HD85                  Q9hd85 homo sapien
   34       8    1.1    731  4   Q96AQ6                  Q96aq6 homo sapien
   35       8    1.1    748  10  Q9FRR2                   Q9frr2 arabidopsis
   36       8    1.1    816  11  O70474                   O70474 rattus norv
   37       8    1.1    858  11  Q8R145                   Q8r145 mus musculu
   38       8    1.1    886  3   O14260                  O14260 schizosacch
   39       8    1.1   1016  16  Q8PKK4                   Q8pkk4 xanthomonas
   40       8    1.1   1290  13  Q9W6E1                   Q9w6e1 gallus gall
   41       8    1.1   1551  5   Q9NGV4                  Q9ngv4 drosophila
   42       8    1.1   1697  12  Q8JJG7                   Q8jjg7 norwalk-lik
   43       8    1.1   1697  12  Q8JJL0                   Q8jjl0 norwalk-lik
   44       8    1.1   2146  5   Q9VC97                  Q9vc97 drosophila
   45       8    1.1   2447  13  O13149                   O13149 fugu rubrip

                               ALIGNMENTS


RESULT 1
O08591
ID   O08591      PRELIMINARY;      PRT;    68 AA.
AC   O08591;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT   "Proteoglycan expression in the normal rat kidney.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U75305; AAB51124.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00054; laminin_G; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
FT   NON_TER       1      1
FT   NON_TER      68     68
SQ   SEQUENCE   68 AA;  7395 MW;  5868E45D8A7083E0 CRC64;

  Query Match             1.8%;  Score 13;  DB 11;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        533 FSRSLPEVPETIE 545
              |||||||||||||
Db         21 FSRSLPEVPETIE 33


RESULT 2
Q9DFL0
ID   Q9DFL0      PRELIMINARY;      PRT;    53 AA.
```

```
AC   Q9DFL0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   S-adenosylmethionine synthase-like protein (Fragment).
OS   Gillichthys mirabilis (Long-jawed mudsucker).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidei;
OC   Gobiidae; Gillichthys.
OX   NCBI_TaxID=8222;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=21117151; PubMed=11172064;
RA   Gracey A.Y., Troll J.V., Somero G.N.;
RT   "Hypoxia-induced gene expression profiling in the euryoxic fish
RT   Gillichthys mirabilis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).
DR   EMBL; AF266225; AAG13345.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   53 AA;  5541 MW;  A7D6F0FF411CF46E CRC64;

  Query Match              1.4%;  Score 10;  DB 13;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 0.053;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        474 QGTRCLCLPG 483
              ||||||||||
Db         35 QGTRCLCLPG 44

RESULT 3
Q80YA8
ID   Q80YA8      PRELIMINARY;      PRT;  1208 AA.
AC   Q80YA8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   5930402A21 protein (Fragment).
GN   5930402A21.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC043114; AAH43114.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 13.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00179; EGF_CA; 13.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 8.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 9.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
FT   NON_TER       1      1
SQ   SEQUENCE   1208 AA;  126978 MW;  FDFF2DF2F0B2F198 CRC64;

  Query Match              1.3%;  Score 9;  DB 11;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        465 EPCLHGGTC 473
              |||||||||
Db        295 EPCLHGGTC 303

RESULT 4
Q7TQ52
ID   Q7TQ52      PRELIMINARY;      PRT;  2516 AA.
AC   Q7TQ52;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Transmembrane receptor Notch1 B.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C.B-17; TISSUE=Thymus;
RA   Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
RT   "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT   the formation of truncated proteins and are involved in the
RT   development of mouse thymic lymphomas.";
RL   Carcinogenesis 24:1-12(2003).
DR   EMBL; AB100603; BAC77038.1; -.
KW   Receptor; Transmembrane.
SQ   SEQUENCE   2516 AA;  269177 MW;  17FD72740EBD6E35 CRC64;

  Query Match              1.3%;  Score 9;  DB 11;  Length 2516;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1014 PCLHGGTCQ 1022

RESULT 5
Q7TQ51
ID   Q7TQ51      PRELIMINARY;      PRT;  2526 AA.
AC   Q7TQ51;
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Transmembrane receptor Notch1 D.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C.B-17; TISSUE=Thymus;
RA   Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
RT   "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT   the formation of truncated proteins and are involved in the
RT   development of mouse thymic lymphomas.";
RL   Carcinogenesis 24:1-12(2003).
DR   EMBL; AB100603; BAC77039.1; -.
KW   Receptor; Transmembrane.
SQ   SEQUENCE   2526 AA;  270583 MW;  017563FCE9703264 CRC64;

  Query Match             1.3%;  Score 9;  DB 11;  Length 2526;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1024 PCLHGGTCQ 1032


RESULT 6
Q8K428
ID   Q8K428      PRELIMINARY;      PRT;  2531 AA.
AC   Q8K428;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Notch 1 protein.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Thymus;
RX   MEDLINE=95044925; PubMed=7956822;
RA   Nye J.S., Kopan R., Axel R.;
RT   "An activated Notch suppresses neurogenesis and myogenesis but not
RT   gliogenesis in mammalian cells.";
RL   Development 120:2421-2430(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Thymus;
RX   MEDLINE=22119593; PubMed=12123574;
RA   Foltz D.R., Santiago M.C., Berechid B.E., Nye J.S.;
RT   "Glycogen Synthase Kinase-3beta Modulates Notch Signaling and
RT   Stability.";
RL   Curr. Biol. 12:1006-1011(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Thymus;
RA   Schellin K.A., Pauley A.M., Nye J.S.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF508809; AAM28905.1; -.
DR   MGD; MGI:97363; Notch1.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0030154; P:cell differentiation; IMP.
DR   GO; GO:0007386; P:compartment specification; IMP.
DR   GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
```

```
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 36.
DR   Pfam; PF00066; notch; 3.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00181; EGF; 37.
DR   SMART; SM00179; EGF_CA; 35.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS00022; EGF_1; 35.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 21.
DR   PIRSF; PIRSF002279; Notch; 1.
KW   ANK repeat; EGF-like domain; Repeat.
SQ   SEQUENCE   2531 AA;  270819 MW;  7DB7E0DEF799D999 CRC64;

  Query Match             1.3%;  Score 9;  DB 11;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 7
Q7TQ50
ID   Q7TQ50      PRELIMINARY;      PRT;  2531 AA.
AC   Q7TQ50;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Transmembrane receptor Notch1.
GN   NOTCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C.B-17; TISSUE=Thymus;
RA   Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
RT   "Radiation-induced deletions in the 5' end region of Notch1 lead to
RT   the formation of truncated proteins and are involved in the
RT   development of mouse thymic lymphomas.";
RL   Carcinogenesis 24:1-12(2003).
DR   EMBL; AB100603; BAC77040.1; -.
KW   Receptor; Transmembrane.
SQ   SEQUENCE   2531 AA;  270832 MW;  97C91F69BABF02BF CRC64;

  Query Match             1.3%;  Score 9;  DB 11;  Length 2531;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        466 PCLHGGTCQ 474
              |||||||||
Db       1029 PCLHGGTCQ 1037


RESULT 8
Q9XZC9
ID   Q9XZC9      PRELIMINARY;      PRT;  3367 AA.
```

```
AC   Q9XZC9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Laminin ALPHA1,2 (SYMBOL=WB).
GN   WB OR WING BLISTER OR CG15288.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=99207061; PubMed=10189378;
RA   Martin D., Zusman S., Li X., Williams E.L., Khare N., DaRocha S.,
RA   Chiquet-Ehrismann R., Baumgartner S.;
RT   "wing blister, a new Drosophila laminin alpha chain required for cell
RT   adhesion and migration during embryonic and imaginal development.";
RL   J. Cell Biol. 145:191-201(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9-Mb region of the genome of
RT   Drosophila melanogaster: the Adh region.";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF135118; AAD31714.1; -.
DR   EMBL; AE003409; AAF44861.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0004002; wb.
DR   GO; GO:0007267; P:cell-cell signaling; NAS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 4.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 16.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
KW   Laminin EGF-like domain.
SQ   SEQUENCE   3367 AA;  374097 MW;  EB125654B1BC1511 CRC64;

  Query Match             1.3%;  Score 9;  DB 5;  Length 3367;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 NDGEWHRVT 611
              |||||||||
Db       3081 NDGEWHRVT 3089

RESULT 9
Q8IP51
ID   Q8IP51      PRELIMINARY;      PRT;  3375 AA.
AC   Q8IP51;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG15288-PB.
GN   WB OR CG15288.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
```

```
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003643; AAN10875.1; -.
DR   FlyBase; FBgn0004002; wb.
DR   GO; GO:0007267; P:cell-cell signaling; NAS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 4.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00180; EGF_Lam; 17.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 4.
SQ   SEQUENCE   3375 AA;  374742 MW;  38CB65C01BB6E416 CRC64;

  Query Match             1.3%;  Score 9;  DB 5;  Length 3375;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        603 NDGEWHRVT 611
              |||||||||
Db       3089 NDGEWHRVT 3097

RESULT 10
Q8I718
ID   Q8I718      PRELIMINARY;      PRT;    97 AA.
AC   Q8I718;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   tRNA intron endonuclease, putative.
GN   PF14_0514.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014824; AAN37127.1; -.
DR   GO; GO:0000214; C:tRNA-intron endonuclease complex; IEA.
DR   GO; GO:0004519; F:endonuclease activity; IEA.
DR   GO; GO:0000213; F:tRNA-intron endonuclease activity; IEA.
DR   GO; GO:0006388; P:tRNA splicing; IEA.
DR   InterPro; IPR006677; tRNA_int_endo_C.
DR   Pfam; PF01974; tRNA_int_endo; 1.
KW   Endonuclease.
SQ   SEQUENCE   97 AA;  11640 MW;  D27B07F8A1360441 CRC64;

  Query Match             1.1%;  Score 8;  DB 5;  Length 97;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        143 EIVFHDLN 150
              ||||||||
Db          9 EIVFHDLN 16

RESULT 11
Q8BMI5
ID   Q8BMI5      PRELIMINARY;      PRT;   219 AA.
AC   Q8BMI5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Weakly similar to neurogenic locus notch 3 protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Forelimb;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK031068; BAC27237.1; -.
DR   PIR; PT0633; PT0633.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
FT   NON_TER       1      1
```

```
SQ   SEQUENCE   219 AA;  23098 MW;  9F2C4BE237F9B522 CRC64;

  Query Match             1.1%;  Score 8;  DB 11;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        482 PGFSGPRC 489
              ||||||||
Db         56 PGFSGPRC 63

RESULT 12
Q9P6X9
ID   Q9P6X9      PRELIMINARY;      PRT;   244 AA.
AC   Q9P6X9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Related to putative lipid binding protein TFS1.
GN   13E11.370.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353820; CAB88605.2; -.
DR   HSSP; P30086; 1BD9.
DR   InterPro; IPR008914; PEBP.
DR   Pfam; PF01161; PBP; 1.
SQ   SEQUENCE   244 AA;  26745 MW;  EEB0DC9E6245012A CRC64;

  Query Match             1.1%;  Score 8;  DB 3;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         90 LAPVNGTS 97
              ||||||||
Db        186 LAPVNGTS 193

RESULT 13
Q82W41
ID   Q82W41      PRELIMINARY;      PRT;   302 AA.
AC   Q82W41;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phosphoadenosine phosphosulfate reductase (EC 2.7.7.4).
GN   CYSD OR NE0856.
OS   Nitrosomonas europaea.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC   Nitrosomonadaceae; Nitrosomonas.
OX   NCBI_TaxID=915;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19718 / IFO 14298;
RX   MEDLINE=22586410; PubMed=12700255;
RA   Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA   Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA   Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT   "Complete genome sequence of the ammonia-oxidizing bacterium and
RT   obligate chemolithoautotroph Nitrosomonas europaea.";
RL   J. Bacteriol. 185:2759-2773(2003).
DR   EMBL; BX321859; CAD84767.1; -.
DR   GO; GO:0016779; F:nucleotidyltransferase activity; IEA.
```

```
DR   GO; GO:0004781; F:sulfate adenylyltransferase (ATP) activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR002500; PAPS_reduct.
DR   Pfam; PF01507; PAPS_reduct; 1.
KW   Nucleotidyltransferase; Transferase; Complete proteome.
SQ   SEQUENCE   302 AA;  34973 MW;  5499EAC71EF9F7DD CRC64;

  Query Match             1.1%;  Score 8;  DB 16;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        279 VLLFSGGK 286
              ||||||||
Db         32 VLLFSGGK 39

RESULT 14
Q842M1
ID   Q842M1      PRELIMINARY;      PRT;   308 AA.
AC   Q842M1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CysD (EC 2.7.7.4).
GN   CYSD.
OS   Rhodococcus sp. DS7.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Nocardiaceae; Rhodococcus.
OX   NCBI_TaxID=227347;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DS7;
RA   Vitale E., Brunelli E., de Ferra F.;
RT   "Effects of cysNC mutations on the desulphurization of organosulphur
RT   compounds in Rhodococcus sp. strain DS7.";
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY244763; AAO89190.1; -.
DR   GO; GO:0016779; F:nucleotidyltransferase activity; IEA.
DR   GO; GO:0004781; F:sulfate adenylyltransferase (ATP) activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR002500; PAPS_reduct.
DR   Pfam; PF01507; PAPS_reduct; 1.
KW   Nucleotidyltransferase; Transferase.
SQ   SEQUENCE   308 AA;  35136 MW;  AE64C09A9D453F7C CRC64;

  Query Match             1.1%;  Score 8;  DB 2;  Length 308;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        279 VLLFSGGK 286
              ||||||||
Db         37 VLLFSGGK 44

RESULT 15
Q8G4F5
ID   Q8G4F5      PRELIMINARY;      PRT;   320 AA.
AC   Q8G4F5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable ribokinase.
GN   BL1431.
OS   Bifidobacterium longum.
OC   Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC   Bifidobacteriaceae; Bifidobacterium.
OX   NCBI_TaxID=216816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCC 2705;
RX   MEDLINE=22294977; PubMed=12381787;
```

```
RA   Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA   Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA   Pridmore R.D., Arigoni F.;
RT   "The genome sequence of Bifidobacterium longum reflects its adaptation
RT   to the human gastrointestinal tract.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR   EMBL; AE014773; AAN25228.1; -.
DR   GO; GO:0004747; F:ribokinase activity; IEA.
DR   GO; GO:0006014; P:D-ribose metabolism; IEA.
DR   InterPro; IPR002173; PfkB.
DR   InterPro; IPR002139; Ribokinase.
DR   Pfam; PF00294; pfkB; 1.
DR   PRINTS; PR00990; RIBOKINASE.
DR   PROSITE; PS00583; PFKB_KINASES_1; 1.
KW   Kinase; Complete proteome.
SQ   SEQUENCE   320 AA;  33130 MW;  28F44619036EFC6B CRC64;

  Query Match             1.1%;  Score 8;  DB 16;  Length 320;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        325 GRWHRVSA 332
              ||||||||
Db        249 GRWHRVSA 256

Search completed: March  9, 2004, 17:30:33
Job time : 64.739 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:15:33 ; Search time 24.1076 Seconds
                                           (without alignments)
                                           2813.016 Million cell updates/sec

Title:          US-10-006-011A-3
Perfect score:  3825
Sequence:       1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_78:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result           Query
  No.    Score   Match Length DB  ID                     Description
-------------------------------------------------------------------------
    1     3825   100.0   4391  2  A38096                 perlecan precursor
    2   3389.5    88.6   3707  2  S18252                 heparan sulfate pr
    3      859    22.5   1959  1  AGRT                   agrin - rat
    4    796.5    20.8   1955  1  AGCH                   agrin precursor -
    5      667    17.4   1328  2  T43060                 agrin - electric r
    6      511    13.4   3375  2  T19821                 hypothetical prote
    7      421    11.0   1715  2  C40228                 neurexin II-alpha
    8    419.5    11.0   1531  2  T42218                 slit-1 protein hom
    9    405.5    10.6   1530  2  I45944                 neurexin I-alpha -
   10    403.5    10.5   1507  2  A40228                 neurexin I-alpha p
   11    387.5    10.1   1025  2  T42626                 secreted leucine-r
   12      387    10.1   1523  2  T13953                 MEGF5 protein - ra
   13      383    10.0   5147  1  IJFFTM                 cadherin-related t
   14      382    10.0   1438  2  A48216                 neurexin III-alpha
   15      382    10.0   1471  2  B48218                 neurexin III-alpha
   16      382    10.0   1578  2  I48216                 neurexin III-alpha
   17    362.5     9.5   4307  2  T20721                 hypothetical prote
   18      357     9.3   1751  1  MMHUMH                 laminin alpha-2 ch
   19    353.5     9.2   3084  1  MMMSA                  laminin alpha-1 ch
   20      353     9.2   3075  2  S14458                 laminin alpha-1 ch
   21      349     9.1   3097  2  T00021                 DN-cadherin - frui
   22      347     9.1   3106  1  S53868                 laminin alpha-2 ch
   23    344.5     9.0    601  2  T22025                 hypothetical prote
   24    344.5     9.0    601  2  D89711                 protein F40E10.4 [
   25      340     8.9   2531  2  T31070                 notch homolog - se
   26    338.5     8.8   1469  2  B36665                 slit protein 2 pre
   27      333     8.7   1480  2  A36665                 slit protein 1 pre
   28      326     8.5   3672  2  T23433                 hypothetical prote
   29      319     8.3   2471  2  A49128                 cell-fate determin
```

```
   30      317     8.3   3704  2  T37316                 probable laminin a
   31      316     8.3   2321  2  S78549                 notch3 protein - h
   32    314.5     8.2   1203  2  A49175                 Motch B protein -
   33      311     8.1   2318  2  S45306                 notch 3 protein -
   34      311     8.1   3034  2  T14119                 seven-pass transme
   35      304     7.9   2809  2  T30213                 G-cadherin - sea u
   36    302.5     7.9   1220  2  A56136                 jagged protein pre
   37    302.5     7.9   2555  2  A40043                 notch protein homo
   38    299.5     7.8   2524  2  A35844                 Xotch protein - Af
   39      298     7.8   4351  2  T00252                 MEGF1 protein - ra
   40      296     7.7   2531  2  S18188                 notch protein homo
   41      296     7.7   2531  2  A46019                 notch-1 protein -
   42      292     7.6    861  2  A48825                 Notch homolog Motc
   43      291     7.6   1964  2  T09059                 notch4 - mouse
   44    287.5     7.5   2352  2  T30201                 Notch homolog prot
   45    284.5     7.4   2703  1  A24420                 notch protein - fr

                              ALIGNMENTS

RESULT 1
A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pro
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R'
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'FL',438-449,'Q',451-502,'A',503-792,'K',794-908,'R'
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515; NID:g29469; PIDN:CAA44373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989
```

```
A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
anes.
A;Reference number: A33625; MUID:90078352; PMID:2687294
A;Accession: B33625
A;Molecule type: protein
A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A;Accession: A33625
A;Molecule type: protein
A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A;Note: peptide potentially matches four different regions of sequence shown
C;Genetics:
A;Gene: GDB:HSPG2
A;Cross-references: GDB:126372; OMIM:142461
A;Map position: 1p36.1-1p36.1
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-4391/Product: perlecan #status predicted <MAT>
F;22-193/Domain: I <DOM1>
F;194-530/Domain: II <DOM2>
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;531-1676/Domain: III <DOM3>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;1677-3686/Domain: IV <DOM4>
F;2007-2034/Domain: transmembrane #status predicted <TRM>
F;3687-4391/Domain: V <DOM5>
F;3845-3880/Domain: EGF homology <EGF1>
F;3888-3921/Domain: EGF homology <EGF>
F;3953-4106/Domain: laminin G repeat homology <LG2>
F;4147-4175/Domain: EGF homology <EGF2>
F;4149-4151/Region: motor neuron attachment (L-R-E) motif
F;4299-4301/Region: motor neuron attachment (L-R-E) motif
F;65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval
F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

  Query Match             100.0%;  Score 3825;  DB 2;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-237;
  Matches  705;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3687 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3746

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3747 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3806

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3807 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3866

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3867 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3926

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3927 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3987 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4046

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4047 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4106

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4107 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4166

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4167 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4226

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4227 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4286

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4287 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 4346

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4347 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
S18252
heparan sulfate proteoglycan - mouse
N;Alternate names: perlecan
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S18252; A31917; B31917; S66460
R;Noonan, D.M.; Fulle, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogl
adhesion molecule.
A;Reference number: S18252; MUID:92078153; PMID:1744087
A;Accession: S18252
A;Molecule type: mRNA
A;Residues: 1-3707 <NOO>
A;Cross-references: EMBL:M77174; NID:g200295; PIDN:AAA39911.1; PID:g200296
R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hass
J. Biol. Chem. 263, 16379-16387, 1988
A;Title: Identification of cDNA clones encoding different domains of the basement membra
A;Reference number: A92680; MUID:89034110; PMID:2972708
A;Accession: A31917
A;Molecule type: mRNA
A;Residues: 940-1601 <NO2>
A;Cross-references: GB:J04054; NID:g200252; PIDN:AAA39899.1; PID:g200253
A;Accession: B31917
A;Molecule type: mRNA
A;Residues: 1870-2600 <NO3>
A;Cross-references: GB:J04055; NID:g200300; PIDN:AAA39912.1; PID:g200301
R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A;Reference number: S66460; MUID:95377282; PMID:7649154
A;Accession: S66460
A;Molecule type: protein
A;Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C;Keywords: glycoprotein
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;764-811/Domain: laminin-type EGF-like homology <LEG>
F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F;3163-3198/Domain: EGF homology <EGF>
F;3270-3423/Domain: laminin G repeat homology <LG2>
F;3464-3492/Domain: EGF homology <EGF7>
F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             88.6%;  Score 3389.5;  DB 2;  Length 3707;
  Best Local Similarity   88.9%;  Pred. No. 3.9e-209;
```

```
  Matches  627;  Conservative   33;  Mismatches   40;  Indels    5;  Gaps    3;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db       3008 EIKITFRPDSADGMLLYNGQKR---SPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3064

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||| ||||||||||||||||||:|||||||||||||||||||||||||||||||
Db       3065 RHPTPLALGQFHTVTLLRSLTQGSLIVGNLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3124

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||:|||||||||||||||||:||| |||||||||:|||||||||| ||||||
Db       3125 IPKAGLSSGFVGCVRELRIQGEEIVFHDVNLTTHGISHCPTCQDRPCQNGGQCQDSESSS 3184

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              | |||||||| :     :     | :   |||||||||||||||||||||||:|||||||
Db       3185 YTCVCPAGFTAAAVNIRKPCTATP-SLWADATCVNRPDGRGYTCRCHLGRSGVRCEEGVT 3243

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              |||||:|||||||||||||||||||||||||||| |:|:|||||||||||||||||||||
Db       3244 VTTPSMSGAGSYLALPALTNTHHELRLDVEFKPLEPNGILLFSGGKSGPVEDFVSLAMVG 3303

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              |||||||||||||||||| ||||||||||||||||||||||||:||||||||||||||||
Db       3304 GHLEFRYELGSGLAVLRSHEPLALGRWHRVSAERLNKDGSLRVDGGRPVLRSSPGKSQGL 3363

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||| ||||||||||| ||||||||||||||||||||||||:|||||||
Db       3364 NLHTLLYLGGVEPSVQLSPATNMSAHFHGCVGEVSVNGKRLDLTYSFLGSQGVGQCYDSS 3423

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              |||||||::||||||||||||||||:||||||||||||||||| ||||:||||:| ||||
Db       3424 PCERQPCRNGATCMPAGEYEFQCLCQDGFKGDLCEHEENPCQLHEPCLNGGTCRGARCLC 3483

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||:|:|: ||||| ||||||||||||||||:|:|||  ||: ||||||||
Db       3484 LPGFSGPRCQQGAGYGVVESDWHPEGSGGNDAPGQYGAYFYDNGFLGLPGNSFSRSLPEV 3543

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||| ||||||| |||||||| | || : ||||||||||||||| |||||||||||| |
Db       3544 PETIEFEVRTSTADGLLLWQGV-VREASRSKDFISLGLQDGHLVFSYQLGSGEARLVSGD 3602

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              |||||||||:||||||:|||||||||:||:||||||||||| |  :|||||||||||| |
Db       3603 PINDGEWHRITALREGQRGSIQVDGEDLVTGRSPGPNVAVNTKDIIYIGGAPDVATLTRG 3662

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              :||||||||:||||||:||||||||||||||||||||||||||||
Db       3663 KFSSGITGCIKNLVLHTARPGAPPPQPLDLQHRAQAGANTRPCPS 3707

RESULT 3
AGRT
agrin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0399; A38856
R;Rupp, F.; Payan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.
Neuron 6, 811-823, 1991
A;Title: Structure and expression of a rat agrin.
A;Reference number: JH0399; MUID:91222570; PMID:1851019
A;Accession: JH0399
A;Molecule type: mRNA
A;Residues: 1-1779;1799-1959 <RUP>
A;Cross-references: GB:M64780; NID:g202798; PIDN:AAA40703.1; PID:g202800
A;Experimental source: embryonic spinal cord
A;Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator
R;Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
J. Neurosci. 12, 3535-3544, 1992
A;Title: Structure and chromosomal localization of the mammalian agrin gene.
A;Reference number: A38856; MUID:92407628; PMID:1326608
A;Accession: A38856
A;Molecule type: mRNA
```

```
A;Residues: 1780-1798 <RU2>
A;Cross-references: GB:S44194
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C;Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.
ylcholine receptor clustering activity.
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-1959/Product: agrin, form 1 #status predicted <AG1>
F;1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
F;1-1779,1799-1959/Product: agrin, form 3 #status predicted <AG3>
F;1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
F;1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
F;22-50/Region: hydrophobic
F;88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
F;163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
F;236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
F;307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
F;381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
F;446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
F;511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
F;540-542/Region: motor neuron attachment (L-R-E) motif
F;596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
F;688-739/Domain: laminin-type EGF-like homology <LE1>
F;742-786/Domain: laminin-type EGF-like homology <LE2>
F;814-864/Domain: Kazal proteinase inhibitor homology <KPI9>
F;869-992/Region: serine/threonine-rich
F;1084-1086/Region: motor neuron attachment (L-R-E) motif
F;1147-1215/Region: serine/threonine-rich
F;1224-1257/Domain: EGF homology <EG1>
F;1287-1442/Domain: laminin G repeat homology <LG1>
F;1444-1476/Domain: EGF homology <EG2>
F;1483-1515/Domain: EGF homology <EG3>
F;1555-1706/Domain: laminin G repeat homology <LG2>
F;1713-1747/Domain: EGF homology <EG4>
F;1807-1959/Domain: laminin G repeat homology <LG3>
F;97-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454-473
-1476,1483-1494,1488-1504,1506-1515/Disulfide bonds: #status predicted
F;145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match            22.5%;  Score 859;  DB 1;  Length 1959;
  Best Local Similarity  31.5%;  Pred. No. 4.6e-47;
  Matches  241;  Conservative   92;  Mismatches  278;  Indels  154;  Gaps   23;

Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              : : ||    :|:|||||             |  ||::  |: || :|||| ||| | :
Db       1287 LALEFRALETEGLLLYNGNA-----------RGKDFLALALLDGRVQFRFDTGSGPAVLT 1335

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYG-- 119
                 |:  | :| : | |   ||:| |    || | |     ||:|:  ||:|| |:
Db       1336 SLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDGLNLDTNLYVGGIPEEQVA 1395

Qy        120 -AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTA---HGISHC--PTCRDRPCQNGGQC 173
                : :  :  |  ||:| | |  :::   |    |    |:  |    |   ||  |  |
Db       1396 MVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVGECGDHPCLPNPCHGGALC 1455

Qy        174 HDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
                 |:  ::| || |  |  |   ::  | |  |   | |     | |  | | |||||
Db       1456 QALEAGMFLCQCPPGRFGPTCADEKS-PCQPNPCHGAAPCRVLSSG-GAKCECPLGRSGT 1513

Qy        234 RCEEGVTVTTPSLSGAG-----------SYLALPALTNTHHEL----RLDVEFKPLAPDG 278
               |:   ||    |  ||           ||| |  |     :|     |:: |   |  |
Db       1514 FCQ---TV----LETAGSRPFLADFNGFSYLELKGLHTFERDLGEKMALEMVFLARGPSG 1566

Qy        279 VLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKD 338
              :||::| |:    ||||||:   |||| |:|| | ||:|| ||:||| | ||  ||  :
Db       1567 LLLYNGQKTDGKGDFVSLALHNRHLEFCYDLGKGAAVIRSKEPIALGTWVRVFLERNGRK 1626

Qy        339 GSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGE 393
              |:|:|  |  ||  || ||:      |||   ||:||      |:    :|: | | :
Db       1627 GALQVGDGPRVLGESP-KSRKVPHTMLNLKEPLYIGGAPDFSKLARGAAVSSGFSGVIQL 1685

Qy        394 VSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDL 453
              ||: | :|      |  : : :  | ||    ||
```

```
Db        1686 VSLRGHQL------LTQEHVLRAVDVSPFADHPCTQA----------------------- 1716

Qy         454 CEHEENPCQLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 509
                        |  |||:||:|   : |  |||  ||||  |::|           :| | |
Db        1717 ---------LGNPCLNGGSCVPREATYECLCPGGFSGLHCEKGL----------VEKSVG 1757

Qy         510 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV-----------PETI-------------- 544
                           |   ||| |  :   |  |           |||:
Db        1758 ------------DLETLAFDGRTYIEYLNAVIESELTNEIPAPETLDSRALFSEKALQSN 1805

Qy         545 --ELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 602
                 || :||    ||:||    :|:| :  |:::| : ||||   | |||    | |   :
Db        1806 HFELSLRTEATQGLVLW----IGKAAERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKV 1861

Qy         603 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGG 660
               |   | |: | || | ||:||  |  |:| ||     ::  |::::||     |
Db        1862 NTNRWLRIRAHREHREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPK 1921

Qy         661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                : :|  ||::::|:          : | |   |      ||||:
Db        1922 AYGTGFVGCLRDVVVGH--------RQLHLLEDAVTKPELRPCPT 1958


RESULT 4
AGCH
agrin precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C;Accession: JH0591; A38857; B38857; I50692
R;Tsim, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
Neuron 8, 677-689, 1992
A;Title: cDNA that encodes active agrin.
A;Reference number: JH0591; MUID:92232297; PMID:1314620
A;Accession: JH0591
A;Molecule type: mRNA
A;Residues: 1-1955 <TSI>
A;Cross-references: GB:M94271; NID:g211120; PIDN:AAA48585.1; PID:g211121
A;Experimental source: brain
R;Ruegg, M.A.; Tsim, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; Gensch, E.M.; McMaha
Neuron 8, 691-699, 1992
A;Title: The agrin gene codes for a family of basal lamina proteins that differ in funct
A;Reference number: A38857; MUID:92232298; PMID:1314621
A;Contents: alternative splicing
A;Accession: A38857
A;Molecule type: mRNA
A;Residues: 1132-1783;1795-1955 <RU2>
A;Cross-references: GB:M97371
A;Accession: B38857
A;Molecule type: mRNA
A;Residues: 1221-1647;1652-1783;1794-1955 <RU3>
A;Cross-references: GB:M97372
A;Note: translation of the nucleotide sequence is not complete
R;Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
Dev. Biol. 158, 523-535, 1993
A;Title: Developmental expression and alternative splicing of chick agrin RNA.
A;Reference number: I50692; MUID:93345745; PMID:8393816
A;Accession: I50692
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'SHLSNEIPA',1784-1795 <THO>
A;Cross-references: EMBL:U07271; NID:g459665; PIDN:AAA16788.1; PID:g459666
C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C;Comment: Alternative splicing produces two inactive proteins: agrin-related protein 1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F;1-38/Domain: signal sequence #status predicted <SIG>
F;39-1955/Product: agrin #status predicted <MAT>
F;39-1783,1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
F;39-1647,1652-1783,1794-1955/Product: agrin-related protein 2 #status predicted <AG2>
F;77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
F;152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
F;225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
F;295-344/Domain: Kazal proteinase inhibitor homology <KPI4>
```

```
F;370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
F;435-483/Domain: Kazal proteinase inhibitor homology <KPI6>
F;500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
F;584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
F;675-726/Domain: laminin-type EGF-like homology <LE1>
F;729-773/Domain: laminin-type EGF-like homology <LE2>
F;801-851/Domain: Kazal proteinase inhibitor homology <KPI9>
F;856-995/Region: serine/threonine-rich
F;1150-1219/Region: serine/threonine-rich
F;1233-1264/Domain: EGF homology <EG1>
F;1294-1448/Domain: laminin G repeat homology <LG1>
F;1429-1431/Region: motor neuron attachment (L-R-E) motif
F;1450-1482/Domain: EGF homology <EG2>
F;1489-1521/Domain: EGF homology <EG3>
F;1560-1711/Domain: laminin G repeat homology <LG2>
F;1718-1751/Domain: EGF homology <EG4>
F;1803-1955/Domain: laminin G repeat homology <LG3>
F;86-105,94-126,160-180,169-201,233-252,241-273,304-323,312-344,378-397,386-418,443-462,
1482,1489-1500,1494-1510,1512-1521/Disulfide bonds: #status predicted
F;390,659,764,814/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             20.8%;  Score 796.5;  DB 1;  Length 1955;
  Best Local Similarity   30.6%;  Pred. No. 4.7e-43;
  Matches  221;  Conservative   99;  Mismatches  271;  Indels  131;  Gaps   25;

Qy           2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
               | : ||     |:||||||           ||  ||||  ||||  | ||: |||   |
Db        1294 IAMEFRATELSGLLLYNGQ-----------NRGKDFISLALVGGFVELRFNTGSGTGVIT 1342

Qy          62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYP-DYGA 120
                   :  | :| : : |:   | | | |   |:| |     ||:|: :|::|| | |  |
Db        1343 SKVRVEPGKWHQLVVNRNRRSGMLAV-DGEHVSGESPTGTDGLNLDTDLFVGGAPEDQMA 1401

Qy         121 I--PKAGLSSGFIGCVR---------ELRIQGEEIVFHDLNLTAHGISHC--PTCRDRPC 167
               :   :   : |  | :|         :|| :| ::::        |:  |    |   ||
Db        1402 VVAERTAATVGLKGSIRLLDVNNQMYDLREKGSDVLY------GSGVGECGNDPCHPNPC 1455

Qy         168 QNGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCH 227
                :|  ||  |:  : | |   :||  |   :   | |  |   |||:  |:| |  | |
Db        1456 HHGASCHVKEAEMFHCECLHSYTGPTCA-DERNPCDPTPCHISATCLVLPEG-GAMCACP 1513

Qy         228 LGRSGLRCEEGVTVT--------TPSLSGAGSYLALPALTN---THHELRLDVEFKPLAP 276
               :|| |  ||    ||         |  :|  ||| |  |     |  :: ::| |   :|
Db        1514 MGREGEFCER---VTEQDHTMPFLPEFNGF-SYLELNGLQTLFLTCRQMSMEVVFLAKSP 1569

Qy         277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLN 336
                |:: ::| |:    ||||||:  |:||:||:|| | ||||| ||: |  |  |  ||
Db        1570 SGMIFYNGQKTDGKGDFVSLALHDGYLEYRYDLGKGAAVLRSKEPVPLNTWISVLLERSG 1629

Qy         337 KDGSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCV 391
               : | :|:| |  |:  || ||:      |||    |:||      |: |  :|  | | |
Db        1630 RKGVMRINNGERVMGESP-KSRKVPHAFLNLKEPFYVGGAPDFSKLARAAAISTSFYGAV 1688

Qy         392 GEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKG 451
                 :|: |  |      |  | |    : |                            |:
Db        1689 QRISIKGVPL------LKEQHIRSAVEIST---------------------------FRA 1715

Qy         452 DLCEHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQ-------GSGHGIA-- 498
                 | :  ||||      :||||    :   | |  ||||  |::       |    ||
Db        1716 HPCTQKPNPCQ------NGGTCSPRLESYECACQRGFSGAHCEKVIIEKAAGDAEAIAFD 1769

Qy         499 ---ESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASG 555
                    ::|   :   ||             | :|     ::|       || ::|    |
Db        1770 GRTYMEYHNAVTKSPDA-------------LDYPAEPSEKALQS--NHFELSIKTEATQG 1814

Qy         556 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 615
               |:||     |:  :  |:|:| : || :   | |||    | |  |||   |  : | |
Db        1815 LILWS----GKGLERSDYIALAIVDGFVQMMYDLGSKPVVLRSTVPINTNHWTHIKAYRV 1870

Qy         616 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNL 673
                | ||:||  |  ::| ||     ::  |::::||     ||      :|:   ||::::
Db        1871 QREGSLQVGNEAPITGSSPLGATQLDTDGALWLGGMERLSVAHKLPKAYSTGFIGCIRDV 1930
```

```
Qy         674 VL 675
               ::
Db        1931 IV 1932

RESULT 5
T43060
agrin - electric ray (Discopyge ommata) (fragment)
C;Species: Discopyge ommata
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
C;Accession: T43060
R;Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMaha
submitted to the EMBL Data Library, September 1992
A;Reference number: Z22308
A;Accession: T43060
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1328 <SMI>
A;Cross-references: EMBL:L01423; NID:g213102; PID:g213103; PIDN:AAA49224.1
C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C;Keywords: glycoprotein; neuromuscular junction

  Query Match             17.4%;  Score 667;  DB 2;  Length 1328;
  Best Local Similarity   27.0%;  Pred. No. 6.2e-35;
  Matches  200;  Conservative  112;  Mismatches  304;  Indels  126;  Gaps   24;

Qy           2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
               | : ||  : ||: |    ::  |    |   :|         |  | ||: |     |
Db         673 ISMEFRASNLDGLPLVQWTEKGKG----LHFYRPS-------EGYVELRFNHGVWDGVIT 721

Qy          62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYG-- 119
                 | :  |::| |   |:   | | |     : | |     ||:|: :|:||| |:
Db         722 SKTLIKPGNWHHVVGNRNRRSGMLSVDGEPHLIGESPPGTDGLNLDTDLFLGGTPEDEMT 781

Qy         120 -AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN------LTAHGISHCPT--CRDRPCQNG 170
                   :   : |  ||:| |:     : :||       |   |:  |    |   ||:|
Db         782 LVTERTTATKGLQGCIRLLDVNN---LIYDLQERSNDVLYGSGVGECGNNPCSPNPCKNR 838

Qy         171 GQCHDSESSSYVCVCPAGFTGSRC--EHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
               |:||  |:  : |     |:|  |  :|:    | |  |   | |:  |:| |  | | :
Db         839 GKCHMKEAEMFHCESVGEFSGPTCADKHNP---CDPNPCHQSANCMVLPEG-GSKCECPM 894

Qy         229 GRSGLRCE-------EGVTVTTPSLSGAGSYLALPAL----TNTHHELRLDVEFKPLAPD 277
               || |  ||       :      |  :|  ||| :  :    ::   :| ::| |    |:
Db         895 GREGELCERVSEAEQDQGKAFIPEFNGL-SYLEMNGIHTFVSDLLQKLSMEVIFLAKDPN 953

Qy         278 GVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNK 337
               |:: ::| |:    ||||| :  |:|||:|:|| | |||||  |: |  |: |: ||  :
Db         954 GMIFYNGQKTDGRGDFVSLNLRDGYLEFKYDLGKGAAVLRSKAPIPLNVWNVVTVERNGR 1013

Qy         338 DGSLRVN------GGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCV 391
                | :::|      |  |  | :|     |||    |:||       : |  : : | | :
Db        1014 KGLMKINKDELVSGESPKSRKAP--HTALNLKEAFYVGGAPDFNKFARAAGIISGFTGAI 1071

Qy         392 GEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKG 451
                ::|:       :   |  : |    : |                            |:
Db        1072 QKLSLK------SIPLLKKENIRNAMEIS--------------------------NFRW 1098

Qy         452 DLCEHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGS 507
                 |    ||||      :|| |    :   |:|  |||||:|::            ||
Db        1099 HACTKTRNPCQ------NGGVCSPRLREYDCMCQRGFSGPQCEKA-----------LE-- 1139

Qy         508 GGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEA 567
                   | |     |:   |: :   |            |: ::|    ||:||     |:
Db        1140 -EKSASGSESVAFNGRTFIEYHNTVTRSEKAVQVNYFEMSIKTEATKGLILWS----GKI 1194

Qy         568 GQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEE 627
                :  |:|:| : || :   | |||    | |  |:|  :| |: | |    |::||  |
Db        1195 AEKSDYIALAVVDGFVQMTYDLGSKPVTLRSTIPVNTNQWVRIKANRIHGYGTLQVGNEA 1254

Qy         628 LVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR----FSSGITGCVKNLVLHSARPGAP 683
                |:| ||     ::  |::::||   :|   | |    :|:|  ||:|::|:
Db        1255 PVTGSSPFAATQLDTDGALWLGGIEKLA--PGNRLPKAYSTGFIGCIKDVVIDR------ 1306
```

```
Qy         684 PPQPLDLQHRAQAGANTRPCPS 705
                 | | |   |        ||:
Db        1307 --QELQLVEDALNNPTILHCPA 1326

RESULT 6
T19821
hypothetical protein ZC101.2e - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Sep-2002
C;Accession: T19821; T19819; T19820; T27490; T27488; T27489; T27487; A47648; B47648; C47
R;Baynes, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19182
A;Accession: T19821
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <WIL>
A;Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone C38C6
A;Accession: T19819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSRKTRTTTKLFGSW
A;Cross-references: EMBL:Z93375; PIDN:CAB07567.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone C38C6
A;Accession: T19820
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694,'H',1883-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPL
A;Cross-references: EMBL:Z93375; PIDN:CAB07568.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone C38C6
R;Percy, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20375
A;Accession: T27490
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3375 <WI2>
A;Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e
A;Experimental source: clone ZC101
A;Accession: T27488
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSRKTRTTTKLFGSW
A;Cross-references: EMBL:Z93395; PIDN:CAB07706.1; GSPDB:GN00020; CESP:ZC101.2a
A;Experimental source: clone ZC101
A;Accession: T27489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1694,'H',1883-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPL
A;Cross-references: EMBL:Z93395; PIDN:CAB07707.1; GSPDB:GN00020; CESP:ZC101.2c
A;Experimental source: clone ZC101
A;Accession: T27487
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1128,1290,'DFARNSPS',1299,'NSS',1303-1304,'R','RHR',1544-1545,'RIRVRS',155
A;Cross-references: EMBL:Z93395; PIDN:CAB07704.1; GSPDB:GN00020; CESP:ZC101.2b
A;Experimental source: clone ZC101
R;Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
Genes Dev. 7, 1471-1484, 1993
A;Title: Products of the unc-52 gene in Caenorhabditis elegans are homologous to the cor
A;Reference number: A47648; MUID:93339574; PMID:8393416
A;Accession: A47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-2441,'R','KRKH',3369,'GN',3372-3373,'G',3375,'RLRHRRRNAQNGPLSR
A;Cross-references: GB:L13458
A;Accession: B47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-2198,'D',2290,'NAR',2294,'L',2296,'WHATE',2302-2303,'V',2305,'
```

```
1,'ANIV',2516-2517,'LQQG',2522,'IDG',2526,'S',2528,'SRGFHV',2535,'F' <RO2>
A;Cross-references: GB:L13458
A;Accession: C47648
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546,'P',548-1128,1290,'DFARNSPS',1299,'NSS',1303-1304,'R','RHR',1544-1545,
A;Cross-references: GB:L13458
C;Genetics:
A;Gene: CESP:ZC101.2e; CESP:ZC101.2a; CESP:ZC101.2c; CESP:ZC101.2b
A;Map position: 2
A;Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2; 1129/1; 1158/3
2; 2613/1; 2684/1; 2757/1; 2813/3; 2863/1; 2900/3; 3084/1; 3176/1; 3250/2
C;Superfamily: LR11 protein; laminin-type EGF-like homology; LDL receptor ligand-binding
C;Keywords: extracellular matrix
F;149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;955-1002/Domain: laminin-type EGF-like homology <LEG1>
F;1011-1058/Domain: laminin-type EGF-like homology <LEG2>

  Query Match             13.4%;  Score 511;  DB 2;  Length 3375;
  Best Local Similarity   20.9%;  Pred. No. 1.8e-24;
  Matches  179;  Conservative  140;  Mismatches  332;  Indels  204;  Gaps   29;

Qy          1 EIKITFRPDSADGMLLYNGQKRVP-----GSPTNLANRQPDFISFGLVGGRPEFRFDAGS 55
              :|:||    :  |::     ||:       :| :  : :     :| |     : || |:
Db       2557 DIEITLNTANPKGIIFET--KRINSGDLLATPYDTIHHEAKITDYGTV----LYEFDIGN 2610

Qy         56 GMATIRHPTPLALGHFHTVTLLRSLTQGSLIVGDLA--------PVNGTSQGKFQGLDLN 107
              |   :    |:    :: : :     | :: : | :        |:   | |      :|
Db       2611 GRQIVETTNPINPNEWNVIKIKNDKNQVTIQLNDESATIRQHTNPLPSLSTG------VN 2664

Qy        108 EELYLGGYPDYGAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPC 167
                :::||  :         :: | | : :: : |  :   |  : :  : :        |
Db       2665 RPVFIGGRHE-----PTNEANDFRGIISQVVLSGHNVGLGDARIPSSVVKYDACASTNLC 2719

Qy        168 QNGGQCHDSES-SSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRC 226
               ||  | :: :   : | |   | |  |:  ::  || |:|     |::  :   : | |
Db       2720 LNGANCRNANNHHGFSCECAEEFHGEYCQW-RSNSCHDESCN-TGICLDNEE--SWQCVC 2775

Qy        227 HLGRSGLRCEEGVTVTTP-SLSGAGSYLALPALTNTHHELRLDVEFKPLAPDG---VLLF 282
               || :||||||   :  |   :   |:||:           : :: :| |      :: |
Db       2776 PLGTTGLRCEEKTEIPQPLGFTSDTSFLAVKRPVKFE---SIKMKLRPQADSDEHILMYF 2832

Qy        283 SGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLR 342
              :       : : ||:::   :         | | |  : :   |   : |  |  | ::|
Db       2833 ASDYGSNTKQYTSLSLIANQVV--------LTVRRPDKEVQKIRSETLEAGEL-IDVAVR 2883

Qy        343 VNGGRPVLRSSPGKSQGLNLHTL-----LYLGGVEPSVPLSPATNMSAHFRGCVGEVSVN 397
                |   |:     :   :   ||     :::||: | :  ||   :   |:||| |: :|
Db       2884 QAGNALVMTVDGNQVSTIETDTLKPGTEIFIGGLPPGLN-SPDDVVEQSFQGCVYEILIN 2942

Qy        398 GKRLDLTYSFLGSQG-IGQCYDS----------------------------------SPC 422
               : :||    | | | |  | :|                                   |
Db       2943 SQDVDL--QNLSSSGDISSCEESQFPVEEDDTTTTTTTEEPEAVIEEPTTEEPTTTEEPI 3000

Qy        423 ERQPCQHGATC----------------------MPAGEYEFQCLCRDG------------ 448
                :| :   |                        |   ||     ||
Db       3001 TEEPTEEPTTTEEPTTTEEPTTTTEEPTTTTTEEPYHIYE---TSRDDDPEIIIPVETTT 3057

Qy        449 -------------------------FKGDLCEHEEN------------------------ 459
                                        : |: : ||
Db       3058 TSTTTTSTTEEPEAEPALVLPTDPVEENDVSDEEEEISTISTVSPDNGLDSDSDYSEGTL 3117

Qy        460 -PCQLREPCLHG---------GTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEG-SG 508
               |    |  : |           |  : |    | :|    :   |   |   : :| :
Db       3118 PPDSSSEEIVVGDVYSTQEPNNICANSTC----GMNGQCVPRNMTHYTCECKLYYDGPTC 3173

Qy        509 GNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAG 568
                  | :: | |  | |:      |     |  | :  : :|   :|:||||| :
Db       3174 SLFKPIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-QRPTVQ 3232

Qy        569 QGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEEL 628
              | :|:||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::|
Db       3233 QMEDYISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRIDNYRE 3292

Qy        629 VSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPL 688
              | ||| |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :        |
Db       3293 VDGRSTGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK--------L 3344

Qy        689 DLQHRAQAGANTRPC 703
              ||   |  | | :||
Db       3345 DLMATAIDGKNVKPC 3359

RESULT 7
C40228
neurexin II-alpha precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Dec-2002
C;Accession: C40228; S27886; S27887
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
Science 257, 50-56, 1992
A;Title: Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin recep
A;Reference number: A40228; MUID:92320296; PMID:1621094
A;Accession: C40228
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1715 <USH>
A;Cross-references: GB:M96376; NID:g205714; PIDN:AAA41707.1; PID:g205716
A;Note: authors translated the codon GAC for residue 1464 as Thr and ACC for residue 146
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1992
A;Description: Neurexins: Synaptic cell surface proteins related to the alpha-lotrotoxin
A;Reference number: S27884
A;Accession: S27886
A;Molecule type: mRNA
A;Residues: 1-1666,'CRK',1670,'PREEKLLPG',1683-1685,'GL',1688,'LDLA',1694-1695,'CCVCRCRA
A;Cross-references: EMBL:M96376; NID:g205714; PIDN:AAA41706.1; PID:g205715
C;Genetics:
A;Introns: 1666/2
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-1715/Product: neurexin II-alpha #status predicted <MAT>
F;694-726/Domain: EGF homology <EGF>
F;1103-1135/Domain: EGF homology <EGF1>

  Query Match             11.0%;  Score 421;  DB 2;  Length 1715;
  Best Local Similarity   21.6%;  Pred. No. 5e-19;
  Matches  193;  Conservative  114;  Mismatches  274;  Indels  312;  Gaps   37;

Qy          2 IKITFRPDSADGMLLYN-GQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              | : ||    :|:||:: |::   |  :: :::: |: :  |: |      | |||   :
Db        517 ISLDFRTTEPNGLLLFSQGRRAGAGVGSHSSSQRADYFAMELLDGYLYLLLDMGSGGIKL 576

Qy         61 RHPT-PLALGHFHTVTLLRSLTQGSLIVGDLA-PVNGTSQGKFQGLDLNEELYLGGYPDY 118
              |  :  :  | :  |   |   :||: |   : |   |  |: : |||  |||||| |:
Db        577 RASSRKVNDGEWCHVDFQRDGRKGSISVNSRSTPFLAT--GESEVLDLESELYLGGLPEG 634

Qy        119 GAIP--------KAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCP--------TC 162
              | :          | | :|::||||:| | |       |      :   |         |
Db        635 GRVDLPLPPEVWTAALRAGYVGCVRDLFIDGRSRDLRGLAEAQGAVGVAPFCSRETLKQC 694

Qy        163 RDRPCQNGGQCHDSESSSYVCVC-PAGFTGSRCE----------------------HSQA 199
                 ||:||| | :   : :|| |   || |  ||                      |::|
Db        695 ASAPCRNGGICREG-WNRFVCDCIGTGFLGRVCEREATVLSYDGSMYMKIMLPNAMHTEA 753

Qy        200 ------------------------------------------LHCHPEACGPDATC---- 213
                                                        | |    | | |
Db        754 EDVSLRFMSQRAYGLMMATTSRESADTLRLELDGGQMKLTVNLDCLRVGCAPSAAAKGPE 813

Qy        214 ------------------VNRP-------------------------------------- 217
                                | |
Db        814 TLFAGHKLNDNEWHTLRVVRRGKSLQLSVDNVTVEGQMAGAHTRLEFHNIETGIMTERRF 873
```

```
Qy        218 -----------------DGRGYTCRCHLG---------RSGLRCEEGVTVTTPSLSGAGS 251
                               :|: |  :|  |         | |||      ||  | |   |
Db        874 ISVVPSNFIGHLSGLVFNGQPYMDQCKDGDITYCELNARFGLRAIVADPVTFKSRS---S 930

Qy        252 YLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGS 311
              ||||  |   :  : |  :||  ||||:|||: |      ||: : :| |:: : ::||:
Db        931 YLALATL-QAYASMHLFFQFKTTAPDGLLLFNSGNG---NDFIVIELVKGYIHYVFDLGN 986

Qy        312 GLAVLR--SAEPLALGRWHRVSAER-LNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYL 368
                ::::  | :|:   :|| |   |      :|:::  | | : | | :: |:|   ||:
Db        987 SPSLMKGNSDKPVNDNQWHNVVVSRDPGNVHTLKID-SRTVTQHSNG-ARNLDLKGELYI 1044

Qy        369 GGVEPSVPLSPATNMSAH---FRGCVGEVSVNGKRLDLTYSFLGSQGIGQCY-----DSS 420
              ||:  ::  |    : |    |:||:  | :||:  ||    |    |||        |:
Db       1045 GGLSKNM-FSNLPKLVASRDGFQGCLASVDLNGRLPDLIADAL--HRIGQVERGCDGPST 1101

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
               |  : | :   |:            |||                      ||  |
Db       1102 TCTEESCANQGVCLQQW---------DGF----------------------TCDCT---- 1126

Qy        481 LPGFSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVF 533
              :  : || |         | |  :    |       || |        |   : |  |
Db       1127 MTSYGGPVCNDPGTTYIFGKGGALITYTW-----PPNDRPSTR----MDRLAVGFSTHQR 1177

Qy        534 SRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGE 593
              |  |          ||  :||||               |:: | :  | :   : :|: :
Db       1178 SAVL----------VRVDSASGL---------------GDYLQLHIDQGTVGVIFNVGTDD 1213

Qy        594 ARLVSEDP---INDGEWHRVTALREGRRGSIQVD----GEELVSGRSPGPNVAV------ 640
                :  ::|   ::||::| |   | |   ::|||     |   :|      :|:
Db       1214 --ITIDEPNAIVSDGKYHVVRFTRSGGNATLQVDSWPVNERYPAGNFDNERLAIARQRIP 1271

Qy        641 ---------------------NAKGSVYIGGAPDVATLTGGRFSSGITGCVKN 672
                                   |:: :: |||        |  |   ::|   |
Db       1272 YRLGRVVDEWLLDKGRQLTIFNSQAAIKIGGRDQ-----GRPFQGQVSGLYYN 1319

RESULT 8
T42218
slit-1 protein homolog - rat
N;Alternate names: MEGF4 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002
C;Accession: T42218
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089; PMID:9693030
A;Accession: T42218
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1531 <NAK>
A;Cross-references: EMBL:AB011530; NID:g3449289; PIDN:BAA32460.1; PID:g3449290
A;Experimental source: strain Sprague-Dawley; brain
C;Genetics:
A;Gene: MEGF4
C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r

  Query Match             11.0%;  Score 419.5;  DB 2;  Length 1531;
  Best Local Similarity   23.3%;  Pred. No. 5.5e-19;
  Matches  173;  Conservative   78;  Mismatches  245;  Indels  245;  Gaps   32;

Qy         25 GSPTNLANRQPDFISFGLVGGR-PEFRF-------------DAGSGMATIRHPTPLALGH 70
              | | |:     |   | || |:   |::              : |    :   | | | :
Db        758 GIPKNVTELYLDGNQFTLVPGQLSTFKYLQLVDLSNNKISSLSNSSFTNMSQLTTLILSY 817

Qy         71 FHTVTL-------LRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIPK 123
                   :       ||||   ||   |   |:   :| |  :     | :|  | |
Db        818 NALQCIPPLAFQGLRSLRLLSLHGND---VSTLQEGIFADVTSLSHLAIGANPLYCDCHL 874

Qy        124 AGLSS---------GFIGCVRELRIQGEEIVFHDLNLTAHG-------------ISHCPT 161
                |||         |   |     ::|:      | ||                 : |
Db        875 RWLSSWVKTGYKEPGIARCAGPPEMEGK------LLLTTPAKKFECQGPPSLAVQAKCDP 928
```

```
Qy        162 CRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATC-VNRPDGR 220
              |   |||| | ||:     | | ||:|: |  || |    |    ||   ||     :
Db        929 CLSSPCQNQGTCHNDPLEVYRCTCPSGYKGRNCEVSLD-SCSSNPCGNGGTCHAQEGEDA 987

Qy        221 GYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVL 280
              |:|| |  |  || |
Db        988 GFTCSCPSGFEGLTC--------------------------------------------- 1002

Qy        281 LFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGS 340
                                                                    :| |
Db       1003 -----------------------------------------------------GMNTDDC 1009

Qy        341 LR---VNGGRPVLRSSPGKSQGLNLHT----LLYLG-GVEPSVPL-SPATNMSAHFRGCV 391
              ::   ||||  |         |:  :|    | | |   |  |   ||  |   |   ||
Db       1010 VKHDCVNGGVCV--------DGIGNYTCQCPLQYTGRACEQLVDFCSPDLNPCQHEAQCV 1061

Qy        392 GEVSVNGKRLDLTYSFLGSQGIGQC-YDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFK 450
              |  :  | | :    : |      |  :   |:   ||:|| |:      : ||| :|:
Db       1062 G--TPEGPRCECVPGYTGD----NCSKNQDDCKDHQCQNGAQCVDEIN-SYACLCAEGYS 1114

Qy        451 GDLCE---HEENPCQLREPCLHGGTC--QGTR--CLCLPGFSGPRCQQGSGHGIAESDWH 503
              | |||      | |:  | :|  |  ||:|  | ||||| || |::       : | :
Db       1115 GQLCEIPPAPRNSCEGTE-CQNGANCVDQGSRPVCQCLPGFGGPECEKLLSVNFVDRDTY 1173

Qy        504 LEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPE-TIELEVRTSTASGLLLWQGV 562
              |:              | |              |   |   | |:| |:  :|:||: |
Db       1174 LQ--------------FTD--------------LQNWPRANITLQVSTAEDNGILLYNG- 1204

Qy        563 EVGEAGQGKDFISLGLQDGHLVFRYQLGS-GEARLVSEDPINDGEWHRVTALREGRRGSI 621
                       | |:: |  ||:   |  ||   : : | : |||::|  |  :   :  ::
Db       1205 -------DNDHIAVELYQGHVRVSYDPGSYPSSAIYSAETINDGQFHTVELVTFDQMVNL 1257

Qy        622 QVDGEELVSGRSPGPNVAVNAKGSVYIGGAP-DVAT--------LTGGRFSSGITGCVKN 672
               :||   :: : | :  :|::  :|:|| | || :        | |  |     ||::|
Db       1258 SIDGGSPMTMDNFGKHYTLNSEAPLYVGGMPVDVNSAAFRLWQILNGTSFH----GCIRN 1313

Qy        673 LVLHS---------ARPGAPP 684
              | :::          :||  |
Db       1314 LYINNELQDFTKTQMKPGVVP 1334

RESULT 9
I45944
neurexin I-alpha - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 09-Dec-2002
C;Accession: I45944
R;Ullrich, B.; Ushkaryov, Y.A.; Sudhof, T.C.
Neuron 14, 497-507, 1995
A;Title: Cartography of neurexins: more than 1000 isoforms generated by alternative spli
A;Reference number: I45944; MUID:95209856; PMID:7695896
A;Accession: I45944
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1530 <ULL>
A;Cross-references: GB:L14855; NID:g388560; PIDN:AAA74123.1; PID:g388561
C;Superfamily: neurexin; EGF homology
F;703-735/Domain: EGF homology <EGF>

  Query Match             10.6%;  Score 405.5;  DB 2;  Length 1530;
  Best Local Similarity   20.5%;  Pred. No. 4.4e-18;
  Matches  203;  Conservative  112;  Mismatches  285;  Indels  391;  Gaps   42;

Qy          2 IKITFRPDSADGMLLY-NGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              |   ||    :|::|: :|: |      :    : || :  :: |      | |||   |
Db        527 ISFDFRTTEPNGLILFSHGKPRHQKDAKHPQMIKVDFFAIEMLDGHLYLLLDMGSGTIKI 586

Qy         61 RHPTPLAL------GHFHTVTLLRSLTQGSLIVGDL-APVNGTSQGKFQGLDLNEELYLG 113
              :     ||      | :: |   |    |:: |  |  |   |: |: | |||::|||||
Db        587 K-----ALQKKVNDGEWYHVDFQRDGRSGTISVNTLRTPY--TAPGESQILDLDDELYLG 639

Qy        114 GYPDYGAIPKAG-----------LSSGFIGCVRELRIQGEEIVFH---DLNLTAHGISHC 159
```

```
               | |:     |||           |: |::||:|:| | |:        ::  ||     |
Db         640 GLPE----NKAGLVFPTEVWTALLNYGYVGCIRDLFIDGQSKDIRQMAEVQSTAGVKPSC 695

Qy         160 -----PTCRDRPCQNGGQCHDSESSSYVCVCP-AGFTGSRCE------------------ 195
                      |   ||:| | | |   : ||| |   |: |  ||
Db         696 SRETAKPCLSNPCKNNGMCRDG-WNRYVCDCSGTGYLGRSCEREATVLSYDGSMFMKIQL 754

Qy         196 ----HSQA------------------------------------------LHCHPEAC-- 207
                   |::|                                          | |    |
Db         755 PVVMHTEAEDVSLRFRSQRAYGILMATTSRDSADTLRLELDAGRVKLTVNLDCIRINCNS 814

Qy         208 --GPDATCV--NRPDGRGYTCR-------------------------------------- 225
                 ||:      |  |   :| |
Db         815 SKGPETLFAGYNLNDNEWHTVRVVRRGKSLKLTVDDQQAMTGQMAGDHTRLEFHNIETGI 874

Qy         226 ---------------------------------------CHL-GRSGLR--CEEGVTVTT 243
                                                      | |  | | |    : ||  |
Db         875 ITERRYLSSVPSNFIGHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIIADPVTFKT 934

Qy         244 PSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHL 303
                |     ||:||  |   :  : |  :||  : ||::|::   ||   ||: : :| |:|
Db         935 KS-----SYVALATL-QAYTSMHLFFQFKTTSLDGLILYN---SGDGNDFIVVELVKGYL 985

Qy         304 EFRYELGSGLAVLR--SAEPLALGRWHRVSAER---------LNKDGSLRVNGGRPVLRS 352
                : ::||:|  :::  | :||   :|| |   |         ::   : ::  |
Db         986 HYVFDLGNGANLIKGSSNKPLNDNQWHNVMISRDTSNLHTVKIDTKITTQITAG------ 1039

Qy         353 SPGKSQGLNLHTLLYLGGV--EPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGS 410
                   :: |:| : ||:|||  |    |    :    |:||:  | :||:  ||
Db        1040 ----ARNLDLKSDLYIGGVAKETYKSLPKLVHAKEGFQGCLASVDLNGRLPDL------- 1088

Qy         411 QGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHG 470
                      |:  |              |: |         :|  ||     ||  : | :
Db        1089 -----ISDALFCN-------------GQIE---------RG--CEGPSTTCQ-EDSCSNQ 1118

Qy         471 GTC----QGTRCLC-LPGFSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGA 518
               | |     |  | | :  |||| |           | |     |       || |
Db        1119 GVCLQQWDGISCDCSMTSFSGPLCNDPGTTYIFSKGGGQITYKW-----PPNDRPSTRA- 1172

Qy         519 YFHDDGFLAFPGHVFSRSLPEVPETIELE---VRTSTASGLLLWQGVEVGEAGQGKDFIS 575
                  |   : |              |:: |   ||  ::|||               |::
Db        1173 ---DRLAIGF--------------STVQKEAVLVRVDSSSGL-------------GDYLE 1202

Qy         576 LGLQDGHLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDG----EELVS 630
               | :  | :  :: :|: :  :   :  ||||::| |   | |   :|||     |   :
Db        1203 LHIHQGKIGVKFNVGTDDIAIEESNAIINDGKYHVVRFTRSGGNATLQVDSWPVIERYPA 1262

Qy         631 GRSPGPNVAV---------------------------NAKGSVYIGG------------- 650
               | :    :|:                           |:: :: |||
Db        1263 GNNDNERLAIARQRIPYRLGRVVDEWLLDKGRQLTIFNSQATIIIGGKEQGQPFQGQLSG 1322

Qy         651 ------------------------------------APDVATLTGGRFSSGITGCVKNLV 674
                                                       ||      |: |      |
Db        1323 LYYNGLKVLNMAAENDANIAIVGNVRLVGEVPSSMTTESTATAMQSEMSTSIMETTTTLA 1382

Qy         675 LHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                 :|| | || :    |        :  |||
Db        1383 TSTARRGKPPTKEPVSQTTDDILVASAECPS 1413

RESULT 10
A40228
neurexin I-alpha precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Dec-2002
C;Accession: A40228; S27884
R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
Science 257, 50-56, 1992
A;Title: Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin recep
A;Reference number: A40228; MUID:92320296; PMID:1621094
A;Accession: A40228
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1507 <USH>
A;Cross-references: GB:M96374; NID:g205710; PIDN:AAA41704.1; PID:g205711
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; transmembrane protein
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1507/Product: neurexin I-alpha #status predicted <MAT>
F;680-712/Domain: EGF homology <EGF>
F;1087-1119/Domain: EGF homology <EGF1>

  Query Match             10.5%;  Score 403.5;  DB 2;  Length 1507;
  Best Local Similarity   20.4%;  Pred. No. 5.8e-18;
  Matches  202;  Conservative  113;  Mismatches  285;  Indels  391;  Gaps   42;

Qy           2 IKITFRPDSADGMLLY-NGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
               |   ||    :|::|: :|: |      :    : || :  :: |      | |||   |
Db         504 ISFDFRTTEPNGLILFSHGKPRHQKDAKHPQMIKVDFFAIEMLDGHLYLLLDMGSGTIKI 563

Qy          61 RHPTPLAL------GHFHTVTLLRSLTQGSLIVGDL-APVNGTSQGKFQGLDLNEELYLG 113
               :     ||      | :: |   |    |:: |  |  |   |: |: : |||::|||||
Db         564 K-----ALQKKVNDGEWYHVDFQRDGRSGTISVNTLRTPY--TAPGESEILDLDDELYLG 616

Qy         114 GYPDYGAIPKAG-----------LSSGFIGCVRELRIQGEEIVFH---DLNLTAHGISHC 159
               | |:     |||           |: |::||:|:| | |:        ::  ||     |
Db         617 GLPE----NKAGLVFPTEVWTALLNYGYVGCIRDLFIDGQSKDIRQMAEIQSTAGVKPSC 672

Qy         160 -----PTCRDRPCQNGGQCHDSESSSYVCVCP-AGFTGSRCE------------------ 195
                      |   ||:| | | |   : ||| |   |: |  ||
Db         673 SRETAKPCLSNPCKNNGMCRDG-WNRYVCDCSGTGYLGRSCEREATVLSYDGSMFMKIQL 731

Qy         196 ----HSQA------------------------------------------LHCHPEAC-- 207
                   |::|                                          | |    |
Db         732 PVVMHTEAEDVSLRFRSQRAYGILMATTSRDSADTLRLELDAGRVKLTVNLDCIRINCNS 791

Qy         208 --GPDATCV--NRPDGRGYTCR-------------------------------------- 225
                 ||:      |  |   :| |
Db         792 SKGPETLFAGYNLNDNEWHTVRVVRRGKSLKLTVDDQQAMTGQMAGDHTRLEFHNIETGI 851

Qy         226 ---------------------------------------CHL-GRSGLR--CEEGVTVTT 243
                                                      | |  | | |    : ||  |
Db         852 ITERRYLSSVPSNFIGHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIIADPVTFKT 911

Qy         244 PSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHL 303
                |     ||:||  |   :  : |  :||  : ||::|::   ||   ||: : :| |:|
Db         912 KS-----SYVALATL-QAYTSMHLFFQFKTTSLDGLILYN---SGDGNDFIVVELVKGYL 962

Qy         304 EFRYELGSGLAVLR--SAEPLALGRWHRVSAER---------LNKDGSLRVNGGRPVLRS 352
                : ::||:|  :::  | :||   :|| |   |         ::   : ::  |
Db         963 HYVFDLGNGANLIKGSSNKPLNDNQWHNVMISRDTSNLHTVKIDTKITTQITAG------ 1016

Qy         353 SPGKSQGLNLHTLLYLGGV--EPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGS 410
                   :: |:| : ||:|||  |    |    :    |:||:  | :||:  ||
Db        1017 ----ARNLDLKSDLYIGGVAKETYKSLPKLVHAKEGFQGCLASVDLNGRLPDL------- 1065

Qy         411 QGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHG 470
                      |:  |              |: |         :|  ||     ||  : | :
Db        1066 -----ISDALFCN-------------GQIE---------RG--CEGPSTTCQ-EDSCSNQ 1095

Qy         471 GTC----QGTRCLC-LPGFSGPRCQQ-------GSGHGIAESDWHLEGSGGNDAPGQYGA 518
               | |     |  | | :  |||| |           | |     |       || |
Db        1096 GVCLQQWDGFSCDCSMTSFSGPLCNDPGTTYIFSKGGGQITHKW-----PPNDRPSTRA- 1149

Qy         519 YFHDDGFLAFPGHVFSRSLPEVPETIELE---VRTSTASGLLLWQGVEVGEAGQGKDFIS 575
                  |   : |              |:: |   ||  ::|||               |::
Db        1150 ---DRLAIGF--------------STVQKEAVLVRVDSSSGL-------------GDYLE 1179

Qy         576 LGLQDGHLVFRYQLGSGEARLVSEDP-INDGEWHRVTALREGRRGSIQVDG----EELVS 630
               | :  | :  :: :|: :  :   :  ||||::| |   | |   :|||     |   :
Db        1180 LHIHQGKIGVKFNVGTDDIAIEESNAIINDGKYHVVRFTRSGGNATLQVDSWPVIERYPA 1239

Qy         631 GRSPGPNVAV---------------------------NAKGSVYIGG------------- 650
               | :    :|:                           |:: :: |||
Db        1240 GNNDNERLAIARQRIPYRLGRVVDEWLLDKGRQLTIFNSQATIIIGGKEQGQPFQGQLSG 1299
```

```
Qy        651 ------------------------------------APDVATLTGGRFSSGITGCVKNLV 674
                                                      ||      |: |      |
Db       1300 LYYNGLKVLNMAAENDANIAIVGNVRLVGEVPSSMTTESTATAMQSEMSTSIMETTTTLA 1359

Qy        675 LHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                :|| | || :    |        :  |||
Db       1360 TSTARRGKPPTKEPISQTTDDILVASAECPS 1390

RESULT 11
T42626
secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment)
N;Alternate names: neurogenic extracellular slit protein
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 16-Aug-2002
C;Accession: T42626
R;Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
Mech. Dev. 79, 57-72, 1998
A;Title: Distinct but overlapping expression patterns of two vertebrate slit homologs i
A;Reference number: Z22177; MUID:99279238; PMID:10349621
A;Accession: T42626
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1025 <HOL>
A;Cross-references: EMBL:AF074960; NID:g4151258; PID:g4151259; PIDN:AAD04345.1
C;Genetics:
A;Gene: Slit2
C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein

  Query Match             10.1%;  Score 387.5;  DB 2;  Length 1025;
  Best Local Similarity   29.7%;  Pred. No. 4e-17;
  Matches  108;  Conservative   55;  Mismatches  144;  Indels   57;  Gaps   18;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |||| |::|  || |  :       |    |
Db        568 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCVCPEGYSGLFCEFSPPMVLPRTSPCDNFDC 626

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   :| |      |:|  |  | :||: |:|   :     ||| :|:      |
Db        627 QNGAQCIIRINEP-----ICQCLPGYLGEKCEKLVSV---NFVNKESYLQIPSAKVRPQT 678

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | ::: :  | :   |: ||  |  : |
Db        679 NITLQIATDED------SGILLYKGDK-----DHIAVELYRGRVRASYDTGSHPASAIYS 727

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |:   || |:|| | : ::  |   ||  : ||:||:
Db        728 VETINDGNFHIVELLTLDSSLSLSVDGGSPKVITNLSKQSTLNFDSPLYVGGMPGKNNVA 787

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : : :   |    || :: | ||  |
Db        788 SLRQAPGQN-GTSFHGCIRNLYINSELQDFRKMPMQTGILPGC---EPCHKKVCAHG-MC 842

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | | :|: | ||:   | || |   |:| |||         | || |  |
Db        843 QPSSQSGFTCECEEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 900

Qy        488 RCQQ 491
               | :
Db        901 LCDE 904

RESULT 12
T13953
MEGF5 protein - rat
N;Alternate names: slit protein homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 16-Aug-2002
C;Accession: T13953
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089; PMID:9693030
A;Accession: T13953
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1523 <NAK>
A;Cross-references: EMBL:AB011531; NID:g3449291; PIDN:BAA32461.1; PID:g3449292
C;Genetics:
A;Gene: MEGF5
C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r

  Query Match             10.1%;  Score 387;  DB 2;  Length 1523;
  Best Local Similarity   28.6%;  Pred. No. 6.7e-17;
  Matches  113;  Conservative   52;  Mismatches  148;  Indels   82;  Gaps   18;

Qy        159 CPT----CRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHS------QALHCHPEACG 208
              | |    |    |::| || |:  : | |:|| ||:|  |||       |   |    |
Db       1071 CETDNDDCVAHKCRHGAQCVDA-VNGYTCICPQGFSGLFCEHPPPMVLLQTSPCDQYECQ 1129

Qy        209 PDATCV---NRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHEL 265
                | |:     |     ||||  | :| |||: :||   :  |  ||: | |      :
Db       1130 NGAQCIVVQQEP-----TCRCPPGFAGPRCEKLITV---NFVGKDSYVEL-ASAKVRPQA 1180

Qy        266 RLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYE-LGSGLAVLRSAEPLAL 324
               : ::      :|:||: |       | ::| :  ||:   |: | |    : | | :
Db       1181 NISLQVATDKDNGILLYKGD-----NDPLALELYQGHVRLVYDSLSSPPTTVYSVETVND 1235

Qy        325 GRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSP----A 380
              |::| |    ||:  :| |: | |       |   : ::: |||||:  |  ||     |
Db       1236 GQFHSVELVMLNQTLNLVVDKGAPKSLGKLQKQPAVGINSPLYLGGIPTSTGLSALRQGA 1295

Qy        381 TNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIG-------------------------- 414
                    | ||: || :| :  |  :  |  | :|
Db       1296 DRPLGGFHGCIHEVRINNELQD--FKALPPQSLGVSPGCKSCTVCRHGLCRSVEKDSVVC 1353

Qy        415 QCYD-----------SSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEE---NP 460
              :|:              ||    | || ||:  |   : | | :|::| ||: :    |
Db       1354 ECHPGWTGPLCDQEAQDPCLGHSCSHG-TCVATGN-SYVCKCAEGYEGPLCDQKNDSANA 1411

Qy        461 CQLREPCLHGGTCQGTR----CLCLPGFSGPRCQQ 491
              |   : | ||      |    ||| |||||  |:|
Db       1412 CSAFK-CHHGQCHISDRGEPYCLCQPGFSGNHCEQ 1445

RESULT 13
IJFFTM
cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
C;Accession: A41087; B41087
R;Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.; Bryant, P.J.; Goodman, C.S.
Cell 67, 853-868, 1991
A;Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe
A;Reference number: A41087; MUID:92069752; PMID:1959133
A;Accession: A41087
A;Molecule type: mRNA
A;Residues: 143-485;1279-5147 <MAH>
A;Cross-references: GB:M80537
A;Accession: B41087
A;Molecule type: DNA
A;Residues: 1-142;487-1278 <MA2>
A;Cross-references: GB:M80537
A;Note: 1229-Gly and 1233-Ser were also found
C;Genetics:
A;Gene: fat
A;Cross-references: FlyBase:FBgn0001075
C;Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
C;Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
F;36-4583/Domain: extracellular #status predicted <EXT>
F;51-156/Domain: cadherin repeat homology <CR1>
F;159-270/Domain: cadherin repeat homology <CR2>
F;271-382/Domain: cadherin repeat homology <CR3>
F;390-494/Domain: cadherin repeat homology <CR4>
```

```
F;497-599/Domain: cadherin repeat homology <CR5>
F;602-708/Domain: cadherin repeat homology <CR6>
F;718-822/Domain: cadherin repeat homology <CR7>
F;831-942/Domain: cadherin repeat homology <CR8>
F;948-1049/Domain: cadherin repeat homology <CR9>
F;1052-1153/Domain: cadherin repeat homology <C10>
F;1156-1278/Domain: cadherin repeat homology <C11>
F;1281-1384/Domain: cadherin repeat homology <C12>
F;1387-1489/Domain: cadherin repeat homology <C13>
F;1492-1601/Domain: cadherin repeat homology <C14>
F;1607-1713/Domain: cadherin repeat homology <C15>
F;1717-1823/Domain: cadherin repeat homology <C16>
F;1826-1922/Domain: cadherin repeat homology <C17>
F;1925-2027/Domain: cadherin repeat homology <C18>
F;2028-2167/Domain: cadherin repeat homology <C19>
F;2169-2278/Domain: cadherin repeat homology <C20>
F;2281-2384/Domain: cadherin repeat homology <C99>
F;2387-2491/Domain: cadherin repeat homology <C21>
F;2494-2596/Domain: cadherin repeat homology <C22>
F;2599-2703/Domain: cadherin repeat homology <C23>
F;2707-2810/Domain: cadherin repeat homology <C24>
F;2813-2913/Domain: cadherin repeat homology <C25>
F;2915-3013/Domain: cadherin repeat homology <C26>
F;3014-3124/Domain: cadherin repeat homology <C27>
F;3127-3229/Domain: cadherin repeat homology <C28>
F;3232-3334/Domain: cadherin repeat homology <C29>
F;3337-3439/Domain: cadherin repeat homology <C30>
F;3442-3545/Domain: cadherin repeat homology <C31>
F;3548-3651/Domain: cadherin repeat homology <C32>
F;3654-3756/Domain: cadherin repeat homology <C33>
F;3954-4010/Domain: EGF homology <EG1>
F;4017-4048/Domain: EGF homology <EG2>
F;4056-4089/Domain: EGF homology <EG3>
F;4096-4127/Domain: EGF homology <EG4>
F;4584-4609/Domain: transmembrane #status predicted <TMM>
F;4610-5147/Domain: intracellular #status predicted <INT>

  Query Match             10.0%;  Score 383;  DB 1;  Length 5147;
  Best Local Similarity   22.9%;  Pred. No. 4.7e-16;
  Matches  185;  Conservative   97;  Mismatches  265;  Indels  260;  Gaps   37;

Qy          4 ITFRPDSADGMLLYNGQKRVPGSPTNL-----------ANRQPDFISFGLVGGRPEFRFD 52
              |  |  : | ::||:  :   |: |||           : :|| ::   |   |  |
Db       3883 IKSRMSNEDEVILYSLLEGGSGNSTNLQLLMAVRLAKTSYQQPKYLIERLREKRSAF--- 3939

Qy         53 AGSGMATIRHPTPLALGHFHTVTLLRSLTQGSLIVG-------DLAPVNGTSQGKFQGLD 105
                                         | |  :|||       |:    |      : ||
Db       3940 -------------------------SELLQKEVIVGYEPCSEPDVCENGGVCSATMRLLD 3974

Qy        106 LNE-------ELYLGG---YPDYGAIPKAGLSSGFIG--CVR--------------ELRI 139
               :         | | |     ||        :||| |  | |              : |
Db       3975 AHSFVIQDSPALVLSGPRVVHDY----SCQCTSGFSGEQCSRRQDPCLPNPCHSQVQCRR 4030

Qy        140 QGEEIVFHDLNLTAHGISHC-----PTCRDRPCQNGGQCHDS-ESSSYVCVCPAGFTGSR 193
               | :  |  :        ||       |  :||:||| |  | : ||| |:|  || |::
Db       4031 LGSD--FQCMCPANRDGKHCEKERSDVCYSKPCRNGGSCQRSPDGSSYFCLCRPGFRGNQ 4088

Qy        194 CEHSQALHCHPEACGPDATCVN-RPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSY 252
              || | :  | |  |     ||: :|   || | |  || |  ||       |      ||
Db       4089 CE-SVSDSCRPNPCLHGGLCVSLKP---GYKCNCTPGRYGRHCERFSYGFQPL-----SY 4139

Qy        253 LALPALTNTHHELRLDVEFKPLAPDGVLLFS-GGKSGPVEDFVSLAMVGGHLEFRYELG- 310
              :  |||  |  :  : : |    |: :||:: | :||   ||::: :|  |    :  |
Db       4140 MTFPALDVTTND--ISIVFATTKPNSLLLYNYGMQSGGRSDFLAIELV--HDRAYFSSGG 4195

Qy        311 --SGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV----NGGRPVLRSSPGKSQ------ 358
                : :: : :   || | ||:|:| |  :  || |    : |       || |
Db       4196 ARTAISTVIAGRNLADGGWHKVTATRNGRVMSLSVAKCADSGDVCTECLPGDSSCYADEV 4255

Qy        359 ----GLNLHTL-LYLGGVEPSVPL--SPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQ 411
                   || :   | :||:  : |:   |    |    ||:  | : |: |:|: | |  :
Db       4256 GPVGTLNFNKQPLMIGGLSSADPILERPGQVHSDDLVGCLHSVHIGGRALNLS-SPLQQK 4314
```

```
Qy        412 GIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGG 471
              ||      : | || ||                                     | |
Db       4315 GI-----LAGCNRQACQ-------------------------------------PALAAE 4332

Qy        472 TCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDA--PGQYGAYFHDDGFLAFP 529
               | |    |:  :|   || |          ||:    :|:  |   |    :  |: |
Db       4333 RCGGFAGQCIDRWSSSLCQCGG---------HLQSPDCSDSLEPITLG----EGAFVEFR 4379

Qy        530 -GHVFSR---------------------------------SLPEVPETIELEVRTSTASG 555
                 :: |                                  : | |: : :  ||    |
Db       4380 ISEIYRRMQLLDNLYNSKSAWLDNQQMRERRAVSNFSTASQIYEAPKMLSMLFRTYKDQG 4439

Qy        556 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSE-DPINDGEWHRVTALR 614
               :|:       |   : | || |::| ||:  :        | |   :|||:|| |:
Db       4440 QILY-------AATNQMFTSLSLREGRLVYYSKQHLTINMTVQETSTLNDGKWHNVSLFS 4492

Qy        615 EGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATL------------TGGRF 662
              | |   : ||| :                    :|   |:| :             ||
Db       4493 ESRSLRLIVDGRQ--------------------VGDELDIAGVHDFLDPYLTILNVGG-- 4530

Qy        663 SSGITGCVKNLVLHSARPGAPPPQPLD 689
                   ||: |: :::        |||:
Db       4531 -EAFVGCLANVTVNNEL------QPLN 4550


RESULT 14
A48216
neurexin III-alpha secreted type 1 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
C;Accession: A48216; B48216
R;Ushkaryov, Y.A.; Suedhof, T.C.
Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
A;Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and
A;Reference number: A48216; MUID:93342001; PMID:8341647
A;Accession: A48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1438 <USH>
A;Cross-references: GB:L14851
A;Accession: B48216
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1368,1372-1438 <US2>
A;Cross-references: GB:L14851
C;Genetics:
A;Introns: 1372/1
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; brain; cell surface component; duplication; extracellu
F;1-27/Domain: signal sequence #status predicted <SIG>
F;202-234/Domain: EGF homology <EGF>
F;651-683/Domain: EGF homology <EGF1>

  Query Match             10.0%;  Score 382;  DB 2;  Length 1438;
  Best Local Similarity   19.8%;  Pred. No. 1.3e-16;
  Matches  175;  Conservative  138;  Mismatches  273;  Indels  296;  Gaps   38;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSG-MAT 59
              || ::|:    :|::|: |             :  |:::  |  |      : |||
Db        286 EITLSFKTWQRNGLILHTG-------------KSADYVNLALKDGAVSLVINLGSGAFEA 332

Qy         60 IRHPTPLALGHF-----HTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGG 114
              |  |     | |     | | : |:| | :: |  :    | :|  :  |  ::  |:||
Db        333 IVEPVN---GKFNDNAWHDVKVTRNLRQVTISVDGILTTTGYTQEDYTMLGSDDFFYVGG 389

Qy        115 YPDYGAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCH 174
               |    :| : :|: |:||::|:  :  :|      |   |
Db        390 SPSTADLPGSPVSNNFMGCLKEVVYKNNDIRLELSRLARIG------------------- 430

Qy        175 DSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLR 234
              |::   |                                               |    :
Db        431 DTKMKIY-----------------------------------------------GEVVFK 443
```

```
Qy         235 CEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVE--- 291
               ||   |:   :     :|::||   ||     :  :|:   |:|::||: ||  | |
Db         444 CENVATLDPINFETPEAYISLPK-WNTKRMGSISFDFRTTEPNGLILFTHGK--PQERKD 500

Qy         292 ---------DFVSLAMVGGHLEFRYELGSGLAVLRSAEPLAL-GRWHRVSAERLNKDGSL 341
                        || :: :: |:|    ::|||   ::: :  |  | |: |  :|  : |::
Db         501 VRSQKNTKVDFFAVELLDGNLYLLLDMGSGTIKVKATQKKANDGEWYHVDIQRDGRSGTI 560

Qy         342 RVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPS-----VPLSPATNMSAH-FRGCVGEVS 395
                ||  |    :| |:|: |:|   :||||:  :     :|    | |  : : ||: ::
Db         561 SVNSRRTPFTAS-GESEILDLEGDMYLGGLPENRAGLILPTELWTAMLNYGYVGCIRDLF 619

Qy         396 VNG--KRLDLTYSFLGSQGI-GQC--YDSSPCERQPCQHGATCMPAGEYEFQCLCR-DGF 449
               ::|  | :        : |:   |    :  |:  ||:: | |   |   | | |   |:
Db         620 IDGRSKNIRQLAEMQNAAGVKSSCSRMSAKQCDSYPCKNNAVCKD-GWNRFICDCTGTGY 678

Qy         450 KGDLCEHEEN--------------PCQLREPC---------------------------- 467
                |  || | :              |  :
Db         679 WGRTCEREASILSYDGSMYMKVIMPMVMHTEAEDVSFRFMSQRAYGLLVATTSRDSADTL 738

Qy         468 ---LHGG--------TCQGTRCLCLPGFSGPRCQQGSGHGIAESDWH------------- 503
                  | ||         |    |       ||     :|  : :::||
Db         739 RLELDGGRVKLMVNLDCIRINC---NSSKGPETLY-AGQKLNDNEWHTVRVVRRGKSLKL 794

Qy         504 ------LEGSGGNDAPGQYGAYFH--DDGFL-----------AFPGHVFS---------- 534
                      ||:   |        ||  : | :           :| ||: |
Db         795 TVDDDVAEGTMVGD---HTRLEFHNIETGIMTEKRYISVVPSSFIGHLQSLMFNGLLYID 851

Qy         535 ------------------RSLPEVPETIE--------------------LEVRTSTASGL 556
                                 |::   | | :                     : :|::| |
Db         852 LCKNGDIDYCELKARFGLRNIIADPVTFKTKSSYLTLATLQAYTSMHLFFQFKTTSADGF 911

Qy         557 LLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV--SEDPINDGEWHRVTALR 614
               :|:       :| | |||:: |  |:: : : ||:    :   |: |:|| :|| |   |
Db         912 ILF------NSGDGNDFIAVELVKGYIHYVFDLGNGPNVIKGNSDRPLNDNQWHNVVITR 965

Qy         615 E-GRRGSIQVDGE---ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSS------ 664
               :     |::|| :   ::::|        :: || :|: |      |  | :|:
Db         966 DNSNTHSLKVDTKVVTQVINGAK-----NLDLKGDLYMAG------LAQGMYSNLPKLVA 1014

Qy         665 ---GITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 703
                  |  ||: :: |:    |  |    |  ||  :|   | |
Db        1015 SRDGFQGCLASVDLN----GRLPDLINDALHR--SGQIDRGC 1050
```

```
RESULT 15
B48218
neurexin III-alpha membrane-bound type 3 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
C;Accession: B48218; C48218
R;Ushkaryov, Y.A.; Suedhof, T.C.
Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
A;Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and
A;Reference number: A48216; MUID:93342001; PMID:8341647
A;Accession: B48218
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1471 <USH>
A;Cross-references: GB:L14851
A;Accession: C48218
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1368,1372-1471 <US2>
A;Cross-references: GB:L14851
C;Genetics:
A;Introns: 1372/1
C;Superfamily: neurexin; EGF homology
C;Keywords: alternative splicing; brain; cell surface component; duplication; receptor;
F;1-27/Domain: signal sequence #status predicted <SIG>
F;202-234/Domain: EGF homology <EGF>
F;651-683/Domain: EGF homology <EGF1>
```

```
  Query Match             10.0%;  Score 382;  DB 2;  Length 1471;
  Best Local Similarity   19.8%;  Pred. No. 1.3e-16;
  Matches  175;  Conservative  138;  Mismatches  273;  Indels  296;  Gaps   38;

Qy           1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSG-MAT 59
               || ::|:    :|::|: |             :  |:::  |  |      : |||
Db         286 EITLSFKTWQRNGLILHTG-------------KSADYVNLALKDGAVSLVINLGSGAFEA 332

Qy          60 IRHPTPLALGHF-----HTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGG 114
               |  |     | |     | | : |:| | :: |  :    | :|  :  |  ::  |:||
Db         333 IVEPVN---GKFNDNAWHDVKVTRNLRQVTISVDGILTTTGYTQEDYTMLGSDDFFYVGG 389

Qy         115 YPDYGAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCH 174
                |    :| : :|: |:||::|:  :  :|      |   |
Db         390 SPSTADLPGSPVSNNFMGCLKEVVYKNNDIRLELSRLARIG------------------- 430

Qy         175 DSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLR 234
               |::   |                                               |    :
Db         431 DTKMKIY-----------------------------------------------GEVVFK 443

Qy         235 CEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVE--- 291
               ||   |:   :     :|::||   ||     :  :|:   |:|::||: ||  | |
Db         444 CENVATLDPINFETPEAYISLPK-WNTKRMGSISFDFRTTEPNGLILFTHGK--PQERKD 500

Qy         292 ---------DFVSLAMVGGHLEFRYELGSGLAVLRSAEPLAL-GRWHRVSAERLNKDGSL 341
                        || :: :: |:|    ::|||   ::: :  |  | |: |  :|  : |::
Db         501 VRSQKNTKVDFFAVELLDGNLYLLLDMGSGTIKVKATQKKANDGEWYHVDIQRDGRSGTI 560

Qy         342 RVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPS-----VPLSPATNMSAH-FRGCVGEVS 395
                ||  |    :| |:|: |:|   :||||:  :     :|    | |  : : ||: ::
Db         561 SVNSRRTPFTAS-GESEILDLEGDMYLGGLPENRAGLILPTELWTAMLNYGYVGCIRDLF 619

Qy         396 VNG--KRLDLTYSFLGSQGI-GQC--YDSSPCERQPCQHGATCMPAGEYEFQCLCR-DGF 449
               ::|  | :        : |:   |    :  |:  ||:: | |   |   | | |   |:
Db         620 IDGRSKNIRQLAEMQNAAGVKSSCSRMSAKQCDSYPCKNNAVCKD-GWNRFICDCTGTGY 678

Qy         450 KGDLCEHEEN--------------PCQLREPC---------------------------- 467
                |  || | :              |  :
Db         679 WGRTCEREASILSYDGSMYMKVIMPMVMHTEAEDVSFRFMSQRAYGLLVATTSRDSADTL 738

Qy         468 ---LHGG--------TCQGTRCLCLPGFSGPRCQQGSGHGIAESDWH------------- 503
                  | ||         |    |       ||     :|  : :::||
Db         739 RLELDGGRVKLMVNLDCIRINC---NSSKGPETLY-AGQKLNDNEWHTVRVVRRGKSLKL 794

Qy         504 ------LEGSGGNDAPGQYGAYFH--DDGFL-----------AFPGHVFS---------- 534
                      ||:   |        ||  : | :           :| ||: |
Db         795 TVDDDVAEGTMVGD---HTRLEFHNIETGIMTEKRYISVVPSSFIGHLQSLMFNGLLYID 851

Qy         535 ------------------RSLPEVPETIE--------------------LEVRTSTASGL 556
                                 |::   | | :                     : :|::| |
Db         852 LCKNGDIDYCELKARFGLRNIIADPVTFKTKSSYLTLATLQAYTSMHLFFQFKTTSADGF 911

Qy         557 LLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV--SEDPINDGEWHRVTALR 614
               :|:       :| | |||:: |  |:: : : ||:    :   |: |:|| :|| |   |
Db         912 ILF------NSGDGNDFIAVELVKGYIHYVFDLGNGPNVIKGNSDRPLNDNQWHNVVITR 965

Qy         615 E-GRRGSIQVDGE---ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSS------ 664
               :     |::|| :   ::::|        :: || :|: |      |  | :|:
Db         966 DNSNTHSLKVDTKVVTQVINGAK-----NLDLKGDLYMAG------LAQGMYSNLPKLVA 1014

Qy         665 ---GITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 703
                  |  ||: :: |:    |  |    |  ||  :|   | |
Db        1015 SRDGFQGCLASVDLN----GRLPDLINDALHR--SGQIDRGC 1050


Search completed: March  9, 2004, 17:22:34
Job time : 28.1076 secs
```

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:19:08 ; Search time 51.7431 Seconds
                                           (without alignments)
                                           2876.963 Million cell updates/sec

Title:          US-10-006-011A-3
Perfect score:  3825
Sequence:       1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters:      809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                         Description
--------------------------------------------------------------------------
    1   901.5    23.6   1931  15  US-10-094-886-180           Sequence 180, App
    2   862.5    22.5   1940  13  US-10-016-283-34            Sequence 34, Appl
    3     840    22.0    775  15  US-10-108-260A-4433         Sequence 4433, Ap
    4     832    21.8    819  10  US-09-866-050A-503          Sequence 503, App
    5   640.5    16.7    492  13  US-10-016-283-36            Sequence 36, Appl
    6     628    16.4    463  15  US-10-104-047-3058          Sequence 3058, Ap
    7     605    15.8    432  9  US-09-764-853-541           Sequence 541, App
    8     605    15.8    432  9  US-09-764-898-179           Sequence 179, App
    9     605    15.8    432  10  US-09-764-881-102           Sequence 102, App
   10     605    15.8    432  14  US-10-073-865-78            Sequence 78, Appl
   11     605    15.8    432  15  US-10-242-747-102           Sequence 102, App
   12   579.5    15.2    456  13  US-10-016-283-25            Sequence 25, Appl
   13     576    15.1    440  13  US-10-016-283-26            Sequence 26, Appl
   14   543.5    14.2    375  9  US-09-978-249-10            Sequence 10, Appl
   15     528    13.8    390  13  US-10-016-283-27            Sequence 27, Appl
   16     441    11.5   3313  9  US-09-737-149-29            Sequence 29, Appl
   17     441    11.5   3313  16  US-10-038-854-69            Sequence 69, Appl
   18   420.5    11.0   1712  15  US-10-291-172-359           Sequence 359, App
   19   419.5    11.0   1531  11  US-09-970-944-31            Sequence 31, Appl
   20   412.5    10.8   1531  11  US-09-970-944-28            Sequence 28, Appl
   21     407    10.6   1534  11  US-09-970-944-30            Sequence 30, Appl
   22     403    10.5   3301  16  US-10-038-854-68            Sequence 68, Appl
   23     398    10.4   1438  15  US-10-297-639-2             Sequence 2, Appli
   24   397.5    10.4   1525  14  US-10-289-776-2             Sequence 2, Appli
   25   397.5    10.4   1529  10  US-09-766-511B-29           Sequence 29, Appl
   26   397.5    10.4   1529  14  US-10-189-123-67            Sequence 67, Appl
   27   397.5    10.4   1529  14  US-10-188-495-67            Sequence 67, Appl
   28   397.5    10.4   3312  14  US-10-225-567A-656          Sequence 656, App
   29   397.5    10.4   3312  16  US-10-038-854-67            Sequence 67, Appl
   30   397.5    10.4   4115  16  US-10-038-854-4             Sequence 4, Appli
   31   394.5    10.3   1523  10  US-09-946-374-198           Sequence 198, App
   32   394.5    10.3   1523  11  US-09-970-944-29            Sequence 29, Appl
   33   394.5    10.3   1523  13  US-10-052-586-290           Sequence 290, App
   34   394.5    10.3   1523  14  US-10-174-590-290           Sequence 290, App
   35   394.5    10.3   1523  14  US-10-176-758-290           Sequence 290, App
   36   394.5    10.3   1523  14  US-10-175-737-290           Sequence 290, App
   37   394.5    10.3   1523  14  US-10-173-706-290           Sequence 290, App
   38   394.5    10.3   1523  14  US-10-175-738-290           Sequence 290, App
   39   394.5    10.3   1523  14  US-10-175-752-290           Sequence 290, App
   40   394.5    10.3   1523  14  US-10-176-482-290           Sequence 290, App
   41   394.5    10.3   1523  14  US-10-176-757-290           Sequence 290, App
   42   394.5    10.3   1523  14  US-10-176-913-290           Sequence 290, App
   43   394.5    10.3   1523  14  US-10-180-552-290           Sequence 290, App
   44   394.5    10.3   1523  14  US-10-180-557-290           Sequence 290, App
   45   394.5    10.3   1523  14  US-10-173-700-290           Sequence 290, App

                              ALIGNMENTS

RESULT 1
US-10-094-886-180
; Sequence 180, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
;  APPLICANT: Kekuda, Ramesh
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Vernet, Corine A.
;  APPLICANT:  Li, Li
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Malyankar, Uriel M.
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Taupier, Raymond J., Jr.
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Pena, Carol
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Gusev, Vladimir
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Gerlach, Valerie
;  APPLICANT:  Pochart, Pascal
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Spaderna, Steven
;  APPLICANT:  LaRochelle, William
;  APPLICANT:  Zhong, Mei
;  TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
;  FILE REFERENCE: 21402-290 B
;  CURRENT APPLICATION NUMBER: US/10/094,886
```

```
;  CURRENT FILING DATE:  2002-03-07
;  PRIOR APPLICATION NUMBER: 60/274,322
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/313,182
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 60/288,052
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: 60/318,510
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/274,281
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/314,018
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: 60/274,194
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/274,849
;  PRIOR FILING DATE: 2001-03-09
;  PRIOR APPLICATION NUMBER: 60/296,693
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: 60/313,626
;  PRIOR FILING DATE: 2001-08-21
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 298
;  SOFTWARE: PatentIn 2.1
; SEQ ID NO 180
;   LENGTH: 1931
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-886-180

  Query Match             23.6%;  Score 901.5;  DB 15;  Length 1931;
  Best Local Similarity   32.2%;  Pred. No. 1.7e-64;
  Matches  242;  Conservative   93;  Mismatches  266;  Indels  151;  Gaps   22;

Qy           2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
               : : ||     |:|||||             |  ||::  |: || : ||| ||| | :
Db        1282 LALEFRALEPQGLLLYNGNA-----------RGKDFLALALLDGRVQLRFDTGSGPAVLT 1330

Qy          62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD---Y 118
                  |:  | :| : | |   :|:| |    || | |     ||:|: :|::|| |:
Db        1331 SAVPVEPGQWHRLELSRHWRRGTLSVDGETPVLGESPSGTDGLNLDTDLFVGGVPEDQAA 1390

Qy         119 GAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNL------TAHGISHC--PTCRDRPCQNG 170
                |: :  : :|  ||:| | :  : :   :| :         |:  |    |   ||  |
Db        1391 VALERTFVGAGLRGCIRLLDVNNQRL---ELGIGPGAATRGSGVGECGDHPCLPNPCHGG 1447

Qy         171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
                 | :|:  : | || |  |  |   ::  | |  |   | |   |:| |  | | |||
Db        1448 APCQNLEAGRFHCQCPPGRVGPTCADEKS-PCQPNPCHGAAPCRVLPEG-GAQCECPLGR 1505

Qy         231 SGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELR--------------LDVEFKPLAP 276
                |  |:     |     |:| :||      :| |||              |:| |    |
Db        1506 EGTFCQ-----TASGQDGSGPFLA-DFNGFSHLELRGLHTFARDLGEKMALEVVFLARGP 1559

Qy         277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLN 336
                |:||::| |:    ||||||:    |||||:|| | ||:|| ||: || | ||| ||
Db        1560 SGLLLYNGQKTDGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNG 1619

Qy         337 KDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSV 396
               : |:|||  |  ||  ||     |||   ||:||      |: |  :|: | | :  ||:
Db        1620 RKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSL 1679

Qy         397 NGKRL----------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQC 443
                |::|          |:| || |           || |    || :||:|:|
Db        1680 GGRQLLTPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP-------- 1720

Qy         444 LCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWH 503
                                   ||            |||  ||||| |::|
Db        1721 --------------------REAAY--------VCLCPGGFSGPHCEKGL---------- 1742

Qy         504 LEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASG 555
               :| | |            |   ||| |  |   |  | |:         || :||    |
Db        1743 VEKSAG------------DVDTLAFDGRTFVEYLNAVTESEKALQSNHFELSLRTEATQG 1790
```

```
Qy         556 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 615
               |:||     |:| :  |:::| : ||||   | |||    | |  |:|   | || | ||
Db        1791 LVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHRE 1846

Qy         616 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNL 673
                | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||::::
Db        1847 QREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDV 1906

Qy         674 VLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
               |:           || |   |      ||||:
Db        1907 VVGR--------HPLHLLEDAVTKPELRPCPT 1930


RESULT 2
US-10-016-283-34
; Sequence 34, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/10/016,283
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US/09/077,955A
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: PCT/US96/20696
;  PRIOR FILING DATE: 1996-12-13
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-10-016-283-34

  Query Match             22.5%;  Score 862.5;  DB 13;  Length 1940;
  Best Local Similarity   32.0%;  Pred. No. 2.8e-61;
  Matches  239;  Conservative   92;  Mismatches  280;  Indels  135;  Gaps   22;

Qy           2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
               : : ||    :|:|||||             |  ||::  |: || :|||| ||| | :
Db        1287 LALEFRALETEGLLLYNGNA-----------RGKDFLALALLDGRVQFRFDTGSGPAVLT 1335

Qy          62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYG-- 119
                  |:  | :| : | |   ||:| |    || | |     ||:|:  ||:|| |:
Db        1336 SLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDGLNLDTNLYVGGIPEEQVA 1395

Qy         120 -AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTA---HGISHC--PTCRDRPCQNGGQC 173
                 : :  :  |  ||:| | |  :::   |    |    |:  |    |   ||  |  |
Db        1396 MVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVGECGDHPCLPNPCHGGALC 1455

Qy         174 HDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
                  |:  ::| || |  |  |   ::  | |  |   | |     | |  | | |||||
Db        1456 QALEAGMFLCQCPPGRFGPTCADEKS-PCQPNPCHGAAPCRVLSSG-GAKCECPLGRSGT 1513

Qy         234 RCEEGVTVTTPSLSGAG-----------SYLALPALTNTHHEL----RLDVEFKPLAPDG 278
                |:   ||    |  ||           ||| |  |     :|     |:: |    | |
Db        1514 FCQ---TV----LETAGSRPFLADFNGFSYLELKGLHTFERDLGEKMALEMVFLARGPSG 1566

Qy         279 VLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKD 338
               :||::| |:    ||||||:   |||| |:|| | ||:|| ||:||| | ||  ||  :
Db        1567 LLLYNGQKTDGKGDFVSLALHNRHLEFCYDLGKGAAVIRSKEPIALGTWVRVFLERNGRK 1626

Qy         339 GSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGE 393
               |:|:|  |  ||  || ||:      |||   ||:||      |:    :|: | | :
Db        1627 GALQVGDGPRVLGESP-KSRKVPHTMLNLKEPLYIGGAPDFSKLARGAAVSSGFSGVIQL 1685

Qy         394 VSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDL 453
               ||: | :|      |  : : :  | ||    ||
Db        1686 VSLRGHQL------LTQEHVLRAVDVSPFADHPCTQA----------------------- 1716

Qy         454 CEHEENPCQLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 509
```

```
                      |  |||:||:|   : |  |||  ||||  |::|           :| | |
Db      1717 ---------LGNPCLNGGSCVPREATYECLCPGGFSGLHCEKGL----------VEKSVG 1757

Qy       510 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASGLLLWQG 561
                         |   ||| |  :   |  | |:         || :||    ||:||
Db      1758 ------------DLETLAFDGRTYIEYLNAVIESEKALQSNHFELSLRTEATQGLVLW-- 1803

Qy       562 VEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSI 621
               :|:| :  |:::| : ||||   | |||    | |   :|   | |: | || | ||:
Db      1804 --IGKAAERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKVNTNRWLRIRAHREHREGSL 1861

Qy       622 QVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNLVLHSAR 679
             ||  |  |:| ||     ::  |::::||     |       : :|  ||::::|:
Db      1862 QVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPKAYGTGFVGCLRDVVVGH-- 1919

Qy       680 PGAPPPQPLDLQHRAQAGANTRPCPS 705
                   : | |   |      ||||:
Db      1920 ------RQLHLLEDAVTKPELRPCPT 1939
```

```
RESULT 3
US-10-108-260A-4433
; Sequence 4433, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4433
;   LENGTH: 775
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-4433

  Query Match             22.0%;  Score 840;  DB 15;  Length 775;
  Best Local Similarity   29.6%;  Pred. No. 5.4e-60;
  Matches  214;  Conservative  107;  Mismatches  257;  Indels  144;  Gaps   20;

Qy         1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
             :| : || :: ||:||| |:           : : ||:|  ::    :|||: |:|:| |
Db       176 QITLEFRAEAEDGLLLYCGENE---------HGRGDFMSLAIIRRSLQFRFNCGTGVAII 226

Qy        61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD-YG 119
                | : || :||| | |    | | : :  || | |||::  :     ||||| |  |
Db       227 VSETKIKLGGWHTVMLYRDGLNGLLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYW 286

Qy       120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN-------LTAHGISHCPT--CRDRPCQNG 170
              :   | : || |||: | : |  |   |:        |:   :  | :  | :  | :|
Db       287 LVRATGTNRGFQGCVQSLAVNGRRI---DMRPWPLGKALSGADVGECSSGICDEASCIHG 343

Qy       171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
             | |   :: ||:|:|| ||
Db       344 GTCTAIKADSYICLCPLGF----------------------------------------- 362

Qy       231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
              |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db       363 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 422

Qy       284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
             | |     ||:|: : |||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db       423 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 477

Qy       344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
             :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db       478 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 537

Qy       404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
              : |  : |:     : || | || || :| |  |  : | |  ||:|
Db       538 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 582
```

```
Qy       464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                 ||   ||      : :| | |                             |  | :
Db       583 ----LH---CQKAIIEAIEIPQFIG---------------------------RSYLTYDN 608

Qy       522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
              |      |   |||       : :  :|:   |||||:|          |||||||:||
Db       609 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 657

Qy       582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
              ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db       658 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 717

Qy       642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
               |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db       718 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 770

Qy       702 PC 703
              |
Db       771 TC 772
```

```
RESULT 4
US-09-866-050A-503
; Sequence 503, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c4U
;  CURRENT APPLICATION NUMBER: US/09/866,050A
;  CURRENT FILING DATE:  2001-05-24
;  NUMBER OF SEQ ID NOS: 725
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 503
;   LENGTH: 819
;   TYPE: PRT
;   ORGANISM: Rat
US-09-866-050A-503

  Query Match             21.8%;  Score 832;  DB 10;  Length 819;
  Best Local Similarity   29.3%;  Pred. No. 2.6e-59;
  Matches  210;  Conservative  109;  Mismatches  264;  Indels  134;  Gaps   17;

Qy         1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
             :| : || :: ||:||| |:           : : ||:|  |:     |||: |:||| |
Db       220 QITLEFRAEAEDGLLLYCGESE---------HGRGDFMSLALIRRSLHFRFNCGTGMAII 270

Qy        61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD-YG 119
                | : || :|:||| |    | | : :  || | |||::  :     ||||| |  |
Db       271 ISETKIKLGAWHSVTLYRDGLNGLLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYW 330

Qy       120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN-------LTAHGISHCPT--CRDRPCQNG 170
              :   | : || |||: | : |::|   |:         |    :  | :  | :  | ||
Db       331 LVRATGTNRGFQGCVQSLAVNGKKI---DMRPWPLGKALNGADVGECSSGICDEASCING 387

Qy       171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
             | |   :: ||:|:|| || |
Db       388 GTCAAIKADSYICLCPLGFRGR-------------------------------------- 409

Qy       231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFSGGK 286
                 ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db       410 ---HCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFTEFEITFRPDSGDGVLLYSYDT 466

Qy       287 SGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGG 346
             |   :||:|: |  ||:|||:: |||  |||| : | ||:|| :   |  |:| |:|:
Db       467 SS--KDFLSIIMAAGHVEFRFDCGSGTGVLRSEDTLTLGQWHDLRVSRTAKNGILQVDKQ 524
```

```
Qy        347 RPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYS 406
              : |   : |    :  :| :::|||     :   : :   | | : :: :| : : : :
Db        525 KVVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGILHPFSGSIQKIILNDRTIHVRHD 584

Qy        407 FLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREP 466
              |  : |:     : ||   || || :| |  |  ::| |  ||:|  |:
Db        585 F--TSGVNVENAAHPCVGAPCAHGGSCRPRKE-GYECDCPLGFEGLNCQKAI-------- 633

Qy        467 CLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFL 526
                       |  : :| | |          | :    : ||  |              |:
Db        634 ---------TEAIEIPQFIGRSYLTYDNPNILK---RVSGSRSN-------------AFM 668

Qy        527 AFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFR 586
               |                    :|:   |||||:|          |||||||:|| |||
Db        669 RF--------------------KTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFS 706

Qy        587 YQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSV 646
              | |||| | ::     :|| |||| |:|:|:| | ||     :|:|||    :|  |::
Db        707 YNLGSGVASIMVNGSFSDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGAL 766

Qy        647 YIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 703
              |:||  ::|  |  ::  |: ||: :  | :          : |   |  | |   |
Db        767 YVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNINTC 816

RESULT 5
US-10-016-283-36
; Sequence 36, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/10/016,283
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US/09/077,955A
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: PCT/US96/20696
;  PRIOR FILING DATE: 1996-12-13
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;   LENGTH: 492
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-016-283-36

  Query Match             16.7%;  Score 640.5;  DB 13;  Length 492;
  Best Local Similarity   33.0%;  Pred. No. 6.6e-44;
  Matches  183;  Conservative   59;  Mismatches  173;  Indels  139;  Gaps   18;

Qy        202 CHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNT 261
              | |  |   | |   |:| |  | | ||| |  |:     |     |:| :||      :
Db         27 CQPNPCHGAAPCRVLPEG-GAQCECPLGREGTFCQ-----TASGQDGSGPFLA-DFNGFS 79

Qy        262 HHELR--------------LDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRY 307
              | |||              |:| |    | |:||::| |:    ||||||:    |||||
Db         80 HLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFRY 139

Qy        308 ELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQG-----LNL 362
              :|| | ||:|| ||: || | ||| ||  : |:|||  |  ||  || ||:      |||
Db        140 DLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESP-KSRKVPHTVLNL 198

Qy        363 HTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL----------DLTYSFLGSQG 412
                 ||:||      |: |  :|: | | :  ||: |::|          |:| || |
Db        199 KEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG--- 254

Qy        413 IGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLH 469
                      || |    || :||:|:|                            ||
Db        255 -------HPCTRASGHPCLNGASCVP----------------------------REAAY- 278

Qy        470 GGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFP 529
                      |||  ||||| |::|           :| | |            |   |||
Db        279 -------VCLCPGGFSGPHCEKGL----------VEKSAG------------DVDTLAFD 309

Qy        530 GHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQGKDF 573
              |  |   |  | |:                 || :||    ||:||     |:| :  |:
Db        310 GRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATERADY 365

Qy        574 ISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRS 633
              ::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:| |
Db        366 VALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSS 425

Qy        634 PGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQ 691
              |     ::  |::::|| |:  |       : :|  ||::::|:           || |
Db        426 PLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLL 477

Qy        692 HRAQAGANTRPCPS 705
                |      ||||:
Db        478 EDAVTKPELRPCPT 491

RESULT 6
US-10-104-047-3058
; Sequence 3058, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE:  2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3058
;   LENGTH: 463
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-104-047-3058

  Query Match             16.4%;  Score 628;  DB 15;  Length 463;
  Best Local Similarity   28.7%;  Pred. No. 6.4e-43;
  Matches  162;  Conservative   80;  Mismatches  198;  Indels  124;  Gaps   16;

Qy        151 LTAHGISHCPT--CRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACG 208
              |:   :  | :  | :  | :|| |   :: ||:|:|| ||
Db         10 LSGADVGECSSGICDEASCIHGGTCTAIKADSYICLCPLGF------------------- 50

Qy        209 PDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL-- 265
                                     |  ||:  |:| |       || | |      | |
Db         51 ----------------------KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSF 88

Qy        266 -RLDVEFKPLAPDGVLLFS---GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEP 321
                 :: |:| : |||||:|   | |     ||:|: : |||:|||:: |||  |||| :|
Db         89 MEFEITFRPDSGDGVLLYSYDTGSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDP 143

Qy        322 LALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPAT 381
              | || || :   |  |:| |:|:  : |   : |    :  :| :::|||     :   :
Db        144 LTLGNWHELRVSRTAKNGILQVDKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNS 203

Qy        382 NMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEF 441
               :   | | : :: :| : : : : |  : |:     : || | || || :| |  |  :
Db        204 GVLKPFSGSIQKIILNDRTIHVKHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GY 260

Qy        442 QCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAE 499
               | |  ||:|                ||   ||      : :| | |
Db        261 DCDCPLGFEG----------------LH---CQKAIIEAIEIPQFIG------------- 288

Qy        500 SDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLW 559
                              |  | : |      |   |||       : :  :|:   |||||
Db        289 --------------RSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLW 325

Qy        560 QGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRG 619
```

```
              :|          |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: |
Db        326 RG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSG 383

Qy        620 SIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSAR 679
               | ||     :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        384 KITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-- 441

Qy        680 PGAPPPQPLDLQHRAQAGANTRPC 703
                      : |   |  | |   |
Db        442 -----DYHISLVEDAVDGKNINTC 460

RESULT 7
US-09-764-853-541
; Sequence 541, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ06
;  CURRENT APPLICATION NUMBER: US/09/764,853
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 939
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 541
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-853-541

  Query Match             15.8%;  Score 605;  DB 9;  Length 432;
  Best Local Similarity   29.0%;  Pred. No. 4.5e-41;
  Matches  157;  Conservative   75;  Mismatches  188;  Indels  122;  Gaps   15;

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   :: ||:|:|| ||
Db          1 GTCTAIKADSYICLCPLGF----------------------------------------- 19

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
               |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db         20 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 79

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: : |||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db         80 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 134

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        135 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 194

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        195 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 239

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        240 ----LH---CQKAIIEAIEIPQFIG---------------------------RSYLTYDN 265

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  :|:   |||||:|          |||||||:||
Db        266 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 314

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        315 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 374

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        375 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 427

Qy        702 PC 703
               |
```

```
Db        428 TC 429

RESULT 8
US-09-764-898-179
; Sequence 179, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ01
;  CURRENT APPLICATION NUMBER: US/09/764,898
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 311
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 179
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-898-179

  Query Match             15.8%;  Score 605;  DB 9;  Length 432;
  Best Local Similarity   29.0%;  Pred. No. 4.5e-41;
  Matches  157;  Conservative   75;  Mismatches  188;  Indels  122;  Gaps   15;

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   :: ||:|:|| ||
Db          1 GTCTAIKADSYICLCPLGF----------------------------------------- 19

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
               |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db         20 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 79

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: : |||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db         80 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 134

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        135 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 194

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        195 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 239

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        240 ----LH---CQKAIIEAIEIPQFIG---------------------------RSYLTYDN 265

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  :|:   |||||:|          |||||||:||
Db        266 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 314

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        315 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 374

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        375 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 427

Qy        702 PC 703
               |
Db        428 TC 429

RESULT 9
US-09-764-881-102
; Sequence 102, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
```

```
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07
;  CURRENT APPLICATION NUMBER: US/09/764,881
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-881-102

  Query Match             15.8%;  Score 605;  DB 10;  Length 432;
  Best Local Similarity   29.0%;  Pred. No. 4.5e-41;
  Matches  157;  Conservative   75;  Mismatches  188;  Indels  122;  Gaps   15;

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   :: ||:|:|| ||
Db          1 GTCTAIKADSYICLCPLGF----------------------------------------- 19

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
               |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db         20 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 79

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: :|||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db         80 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 134

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        135 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 194

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        195 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 239

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        240 ----LH---CQKAIIEAIEIPQFIG--------------------------RSYLTYDN 265

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  :|:   |||||:|          |||||||:||
Db        266 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 314

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        315 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 374

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        375 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 427

Qy        702 PC 703
               |
Db        428 TC 429

RESULT 10
US-10-073-865-78
; Sequence 78, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ09C1
;  CURRENT APPLICATION NUMBER: US/10/073,865
;  CURRENT FILING DATE:  2002-02-14
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 154
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
;   LENGTH: 432
```

```
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-073-865-78

  Query Match             15.8%;  Score 605;  DB 14;  Length 432;
  Best Local Similarity   29.0%;  Pred. No. 4.5e-41;
  Matches  157;  Conservative   75;  Mismatches  188;  Indels  122;  Gaps   15;

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   :: ||:|:|| ||
Db          1 GTCTAIKADSYICLCPLGF----------------------------------------- 19

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
               |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db         20 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 79

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: :|||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db         80 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 134

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        135 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 194

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        195 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 239

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        240 ----LH---CQKAIIEAIEIPQFIG--------------------------RSYLTYDN 265

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  :|:   |||||:|          |||||||:||
Db        266 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 314

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        315 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 374

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        375 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 427

Qy        702 PC 703
               |
Db        428 TC 429

RESULT 11
US-10-242-747-102
; Sequence 102, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ07C1
;  CURRENT APPLICATION NUMBER: US/10/242,747
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: 09/764,881
;  PRIOR FILING DATE: 2001-01-17
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: 60/214,886
;  PRIOR FILING DATE: 2000-06-28
;  PRIOR APPLICATION NUMBER: 60/217,487
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,758
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/220,963
;  PRIOR FILING DATE: 2000-07-26
```

```
;  PRIOR APPLICATION NUMBER: 60/217,496
;  PRIOR FILING DATE: 2000-07-11
;  PRIOR APPLICATION NUMBER: 60/225,447
;  PRIOR FILING DATE: 2000-08-14
;  PRIOR APPLICATION NUMBER: 60/218,290
;  PRIOR FILING DATE: 2000-07-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 192
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 102
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-242-747-102

  Query Match             15.8%;  Score 605;  DB 15;  Length 432;
  Best Local Similarity   29.0%;  Pred. No. 4.5e-41;
  Matches  157;  Conservative   75;  Mismatches  188;  Indels  122;  Gaps   15;

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   ::  ||:|:|| ||
Db          1 GTCTAIKADSYICLCPLGF----------------------------------------- 19

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
               |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db         20 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 79

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: : |||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db         80 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 134

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        135 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 194

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        195 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 239

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        240 ----LH---CQKAIIEAIEIPQFIG-----------------------------RSYLTYDN 265

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  |:|   |||||:|          |||||||:||
Db        266 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 314

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        315 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 374

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        375 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 427

Qy        702 PC 703
               |
Db        428 TC 429

RESULT 12
US-10-016-283-25
; Sequence 25, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/10/016,283
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US/09/077,955A
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: PCT/US96/20696
;  PRIOR FILING DATE: 1996-12-13
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-016-283-25

  Query Match             15.2%;  Score 579.5;  DB 13;  Length 456;
  Best Local Similarity   32.7%;  Pred. No. 6e-39;
  Matches  168;  Conservative   57;  Mismatches  156;  Indels  133;  Gaps   16;

Qy        242 TTPSLSGAGSYLALPALTNTHHELR--------------LDVEFKPLAPDGVLLFSGGKS 287
              |     |:| :||      :| |||              |:| |    | |:||::| |:
Db         25 TASGQDGSGPFLA-DFNGFSHLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKT 83

Qy        288 GPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGR 347
                  ||||||:    |||||:|| | ||:|| ||: || | ||| ||  : |:|||  |
Db         84 DGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGP 143

Qy        348 PVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL- 401
               ||  || ||:      |||   ||:||      |: |  :|: | | :  ||: |::|
Db        144 RVLGESP-KSRKVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLL 202

Qy        402 ---------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGF 449
                       |:| || |           || |    || :||:|:|
Db        203 TPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP-------------- 237

Qy        450 KGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 509
                            ||            |||  ||||| |::|           :| | |
Db        238 --------------REAAY--------VCLCPGGFSGPHCEKGL----------VEKSAG 265

Qy        510 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVRTSTA 553
                          |   ||| |  |   |  | |:                 || :||
Db        266 ------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEAT 313

Qy        554 SGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTAL 613
               ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | || |
Db        314 QGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAH 369

Qy        614 REGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVK 671
              || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||::
Db        370 REQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLR 429

Qy        672 NLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              ::|:           || |   |      ||||:
Db        430 DVVVGR--------HPLHLLEDAVTKPELRPCPT 455

RESULT 13
US-10-016-283-26
; Sequence 26, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/10/016,283
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US/09/077,955A
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: PCT/US96/20696
;  PRIOR FILING DATE: 1996-12-13
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 440
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-016-283-26

  Query Match             15.1%;  Score 576;  DB 13;  Length 440;
```

```
  Best Local Similarity   32.9%;  Pred. No. 1.1e-38;
  Matches  163;  Conservative   55;  Mismatches  145;  Indels  132;  Gaps   15;

Qy       261 THHELR--------------LDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFR 306
             :| |||              |:| |    | |:||::| |:    ||||||:    ||||
Db        27 SHLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFR 86

Qy       307 YELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQG-----LN 361
             |:|| | ||:|| ||: || | ||| ||  : |:|||  |  ||  || ||:      ||
Db        87 YDLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESP-KSRKVPHTVLN 145

Qy       362 LHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL----------DLTYSFLGSQ 411
             |   ||:||      |: |  :|: | |:   ||: |::|          |:| || |
Db       146 LKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG-- 202

Qy       412 GIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCL 468
                      || |    || :||:|:|                            ||
Db       203 --------HPCTRASGHPCLNGASCVP----------------------------REAAY 226

Qy       469 HGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAF 528
                      |||  ||||| |::|           :| | |            |   |||
Db       227 --------VCLCPGGFSGPHCEKGL----------VEKSAG------------DVDTLAF 256

Qy       529 PGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQGKD 572
              |  |   |  | |:                 || :||    ||:||     |:| :  |
Db       257 DGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATERAD 312

Qy       573 FISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGR 632
             :::| : ||||   | |||    | | :| |   | || | || | ||:||  |  |:|
Db       313 YVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGS 372

Qy       633 SPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 690
             ||     ::  |::::|| |:  |       : :|  ||::::|:           || |
Db       373 SPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHL 424

Qy       691 QHRAQAGANTRPCPS 705
                |      ||||:
Db       425 LEDAVTKPELRPCPT 439
```

```
RESULT 14
US-09-978-249-10
; Sequence 10, Application US/09978249
; Patent No. US20020106780A1
; GENERAL INFORMATION:
;  APPLICANT: Fiscella, et al.
;  TITLE OF INVENTION: Extracellular Matrix Polynucleotides, Polypeptides, and Antibodie
;  FILE REFERENCE: PT054P1
;  CURRENT APPLICATION NUMBER: US/09/978,249
;  CURRENT FILING DATE:  2001-10-17
;  PRIOR APPLICATION NUMBER: PCT/US01/11643
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: 60/198,123
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 375
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-978-249-10

  Query Match             14.2%;  Score 543.5;  DB 9;  Length 375;
  Best Local Similarity   30.6%;  Pred. No. 4.1e-36;
  Matches  136;  Conservative   70;  Mismatches  161;  Indels   77;  Gaps   12;

Qy       265 LRLDVEFKPLAPDGVLLFS---GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEP 321
             :  :: |:| : |||||:|   | |     ||:|: : |||:|||:: |||  |||| :|
Db         1 MEFEITFRPDSGDGVLLYSYDTGSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDP 55

Qy       322 LALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPAT 381
             | || || :   |  |:| |:|:  : |     |    :  :| :::|||     :   :
Db        56 LTLGNWHELRVSRTAKNGILQVDKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNS 115
```

```
Qy       382 NMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEF 441
              :   | | : :: :| : : : : |  : |:     : || | || || :| |  |  :
Db       116 GVLKPFSGSIQKIILNDRTIHVKHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GY 172

Qy       442 QCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAE 499
              | |  ||:|                ||   ||      : :| | |
Db       173 DCDCPLGFEG----------------LH---CQKAIIEAIEIPQFIG------------- 200

Qy       500 SDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLW 559
                             |  | : |      |   |||       : :  :|:   |||||
Db       201 --------------RSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLW 237

Qy       560 QGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRG 619
             :|          |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: |
Db       238 RG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSG 295

Qy       620 SIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSAR 679
              | ||     :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db       296 KITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-- 353

Qy       680 PGAPPPQPLDLQHRAQAGANTRPC 703
                     : |   |  | |   |
Db       354 -----DYHISLVEDAVDGKNINTC 372
```

```
RESULT 15
US-10-016-283-27
; Sequence 27, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/10/016,283
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US/09/077,955A
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: PCT/US96/20696
;  PRIOR FILING DATE: 1996-12-13
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
;   LENGTH: 390
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-016-283-27

  Query Match             13.8%;  Score 528;  DB 13;  Length 390;
  Best Local Similarity   33.0%;  Pred. No. 8.1e-35;
  Matches  149;  Conservative   50;  Mismatches  134;  Indels  118;  Gaps   14;

Qy       291 EDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVL 350
             :| ||||:    |||||:|| | ||:|| ||: || | ||| ||  : |:|||  |  ||
Db        21 DDKVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVL 80

Qy       351 RSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL---- 401
               || ||:      |||   ||:||      |: |  :|: | | :  ||: |::|
Db        81 GESP-KSRKVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPE 139

Qy       402 ------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGD 452
                   |:| || |           || |    || :||:|:|
Db       140 HVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP----------------- 171

Qy       453 LCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDA 512
                        ||            |||  ||||| |::|           :| | |
Db       172 -----------REAAY--------VCLCPGGFSGPHCEKGL----------VEKSAG--- 199

Qy       513 PGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVRTSTASGL 556
                      |   ||| |  |   |  | |:                 || :||    ||
Db       200 ---------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGL 250

Qy       557 LLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREG 616
```

```
                :||     |:| :  |:::| : ||||   | |||    | |  |:|   | || | ||
Db          251 VLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQ 306

Qy          617 RRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLV 674
                | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||::::|
Db          307 REGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVV 366

Qy          675 LHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
                :           || |   |      ||||:
Db          367 VGR--------HPLHLLEDAVTKPELRPCPT 389

Search completed: March  9, 2004, 17:25:14
Job time : 54.7431 secs
```

```
                   GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:16:23 ; Search time 27.6355 Seconds
                                           (without alignments)
                                           1317.011 Million cell updates/sec

Title:          US-10-006-011A-3
Perfect score:  3825
Sequence:       1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                %
Result          Query
   No.   Score  Match Length DB  ID                        Description
-------------------------------------------------------------------------
     1   862.5   22.5   1940  2  US-08-644-271-30           Sequence 30, Appl
     2   862.5   22.5   1940  4  US-09-077-955-34           Sequence 34, Appl
     3   640.5   16.7    492  2  US-08-644-271-32           Sequence 32, Appl
     4   640.5   16.7    492  4  US-09-077-955-36           Sequence 36, Appl
     5   579.5   15.2    456  4  US-09-077-955-25           Sequence 25, Appl
     6     576   15.1    440  4  US-09-077-955-26           Sequence 26, Appl
     7     528   13.8    390  4  US-09-077-955-27           Sequence 27, Appl
     8   397.5   10.4    771  3  US-09-188-930-183          Sequence 183, App
     9   397.5   10.4   1525  3  US-09-191-647-2            Sequence 2, Appli
    10   397.5   10.4   1525  3  US-09-540-245A-2           Sequence 2, Appli
    11   397.5   10.4   1525  3  US-09-540-153-2            Sequence 2, Appli
    12   394.5   10.3   1529  4  US-09-312-283C-396         Sequence 396, App
    13     394   10.3    716  4  US-09-312-283C-183         Sequence 183, App
    14   391.5   10.2   1523  4  US-09-182-024A-2           Sequence 2, Appli
    15     382   10.0    338  4  US-09-077-955-28           Sequence 28, Appl
    16     361    9.4   3075  2  US-08-460-309-5            Sequence 5, Appli
    17     361    9.4   3075  2  US-08-125-077-5            Sequence 5, Appli
    18     359    9.4   1130  6  5444158-2                 Patent No. 5444158
    19     357    9.3   1130  2  US-08-460-309-2            Sequence 2, Appli
    20     357    9.3   1130  2  US-08-125-077-2            Sequence 2, Appli
    21     357    9.3   3088  4  US-09-562-702A-8           Sequence 8, Appli
    22     357    9.3   3089  4  US-09-562-702A-4           Sequence 4, Appli
    23     357    9.3   3110  4  US-09-562-702A-2           Sequence 2, Appli
    24     357    9.3   3110  4  US-09-562-702A-6           Sequence 6, Appli
    25     357    9.3   3110  4  US-09-561-709B-7           Sequence 7, Appli
    26     357    9.3   3111  2  US-08-460-309-4            Sequence 4, Appli
    27     357    9.3   3111  2  US-08-125-077-4            Sequence 4, Appli
    28   349.5    9.1    735  3  US-09-191-647-9            Sequence 9, Appli
    29   349.5    9.1    735  3  US-09-540-245A-9           Sequence 9, Appli
    30   349.5    9.1    735  3  US-09-540-153-9            Sequence 9, Appli
    31     347    9.1   3084  4  US-09-562-702A-12          Sequence 12, Appl
    32     347    9.1   3106  4  US-09-562-702A-10          Sequence 10, Appl
    33     340    8.9    294  4  US-09-077-955-29           Sequence 29, Appl
    34     333    8.7   1480  3  US-09-191-647-7            Sequence 7, Appli
    35     333    8.7   1480  3  US-09-540-245A-7           Sequence 7, Appli
    36     333    8.7   1480  3  US-09-540-153-7            Sequence 7, Appli
    37     333    8.7   1480  5  PCT-US91-09055-2           Sequence 2, Appli
    38   328.5    8.6   2471  1  US-08-185-432-16           Sequence 16, Appl
    39   328.5    8.6   2471  1  US-08-083-590A-19          Sequence 19, Appl
    40   328.5    8.6   2471  3  US-08-532-384-19           Sequence 19, Appl
    41   328.5    8.6   2471  4  US-08-899-232-1            Sequence 1, Appli
    42     328    8.6   1480  4  US-09-182-024A-5           Sequence 5, Appli
    43     326    8.5    256  4  US-09-077-955-30           Sequence 30, Appl
    44   322.5    8.4   3571  4  US-09-911-842A-2           Sequence 2, Appli
    45     316    8.3   2321  4  US-09-230-652-2            Sequence 2, Appli

                                 ALIGNMENTS


RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1940 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Rat Agrin
;      LOCATION:  1...1940
;      OTHER INFORMATION:
US-08-644-271-30

  Query Match             22.5%;  Score 862.5;  DB 2;  Length 1940;
```

```
  Best Local Similarity   32.0%;  Pred. No. 7.4e-64;
  Matches  239;  Conservative   92;  Mismatches  280;  Indels  135;  Gaps   22;

Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              : : ||    :|:|||||             |  ||::  |: || :|||| ||| | :
Db       1287 LALEFRALETEGLLLYNGNA-----------RGKDFLALALLDGRVQFRFDTGSGPAVLT 1335

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYG-- 119
                 |:  | :| : | |   ||:| |    || | |     ||:|:  ||:|| |:
Db       1336 SLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDGLNLDTNLYVGGIPEEQVA 1395

Qy        120 -AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTA---HGISHC--PTCRDRPCQNGGQC 173
                : :  :  |  ||:| | |  :::   |    |    |:  |    |   ||  |  |
Db       1396 MVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVGECGDHPCLPNPCHGGALC 1455

Qy        174 HDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
                 |:  ::| || |  |  |   ::  | |  |   | |     | |  | | |||||
Db       1456 QALEAGMFLCQCPPGRFGPTCADEKS-PCQPNPCHGAAPCRVLSSG-GAKCECPLGRSGT 1513

Qy        234 RCEEGVTVTTPSLSGAG-----------SYLALPALTNTHHEL----RLDVEFKPLAPDG 278
               |:   ||    |  ||           ||| |  |     :|     |:: |    | |
Db       1514 FCQ---TV----LETAGSRPFLADFNGFSYLELKGLHTFERDLGEKMALEMVFLARGPSG 1566

Qy        279 VLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKD 338
              :||::| |:    ||||||:   |||| |:|| | ||:|| ||:||| | ||  ||  :
Db       1567 LLLYNGQKTDGKGDFVSLALHNRHLEFCYDLGKGAAVIRSKEPIALGTWVRVFLERNGRK 1626

Qy        339 GSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGE 393
              |:|:|  |  ||  || ||:      |||   ||:||      |:    :|: | | :
Db       1627 GALQVGDGPRVLGESP-KSRKVPHTMLNLKEPLYIGGAPDFSKLARGAAVSSGFSGVIQL 1685

Qy        394 VSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDL 453
              ||: | :|      |  : : :  | ||    ||
Db       1686 VSLRGHQL------LTQEHVLRAVDVSPFADHPCTQA----------------------- 1716

Qy        454 CEHEENPCQLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 509
                       |  |||:||:|   : |  |||  ||||  |::|           :| | |
Db       1717 ---------LGNPCLNGGSCVPREATYECLCPGGFSGLHCEKGL----------VEKSVG 1757

Qy        510 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASGLLLWQG 561
                          |   ||| |  :   |  | |:         || :||    ||:||
Db       1758 ------------DLETLAFDGRTYIEYLNAVIESEKALQSNHFELSLRTEATQGLVLW-- 1803

Qy        562 VEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSI 621
                :|:| :  |:::| : ||||   | |||    | |   :|   | |: | || | ||:
Db       1804 --IGKAAERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKVNTNRWLRIRAHREHREGSL 1861

Qy        622 QVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNLVLHSAR 679
              ||  |  |:| ||     ::  |::::||     |       : :|  ||::::|:
Db       1862 QVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPKAYGTGFVGCLRDVVVGH-- 1919

Qy        680 PGAPPPQPLDLQHRAQAGANTRPCPS 705
                    : | |   |      ||||:
Db       1920 ------RQLHLLEDAVTKPELRPCPT 1939

RESULT 2
US-09-077-955-34
; Sequence 34, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 1940
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-077-955-34

  Query Match             22.5%;  Score 862.5;  DB 4;  Length 1940;
  Best Local Similarity   32.0%;  Pred. No. 7.4e-64;
  Matches  239;  Conservative   92;  Mismatches  280;  Indels  135;  Gaps   22;

Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              : : ||    :|:|||||             |  ||::  |: || :|||| ||| | :
Db       1287 LALEFRALETEGLLLYNGNA-----------RGKDFLALALLDGRVQFRFDTGSGPAVLT 1335

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYG-- 119
                 |:  | :| : | |   ||:| |    || | |     ||:|:  ||:|| |:
Db       1336 SLVPVEPGRWHRLELSRHWRQGTLSVDGETPVVGESPSGTDGLNLDTNLYVGGIPEEQVA 1395

Qy        120 -AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTA---HGISHC--PTCRDRPCQNGGQC 173
                : :  :  |  ||:| | |  :::   |    |    |:  |    |   ||  |  |
Db       1396 MVLDRTSVGVGLKGCIRMLDINNQQLELSDWQRAAVQSSGVGECGDHPCLPNPCHGGALC 1455

Qy        174 HDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
                 |:  ::| || |  |  |   ::  | |  |   | |     | |  | | |||||
Db       1456 QALEAGMFLCQCPPGRFGPTCADEKS-PCQPNPCHGAAPCRVLSSG-GAKCECPLGRSGT 1513

Qy        234 RCEEGVTVTTPSLSGAG-----------SYLALPALTNTHHEL----RLDVEFKPLAPDG 278
               |:   ||    |  ||           ||| |  |     :|     |:: |    | |
Db       1514 FCQ---TV----LETAGSRPFLADFNGFSYLELKGLHTFERDLGEKMALEMVFLARGPSG 1566

Qy        279 VLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKD 338
              :||::| |:    ||||||:   |||| |:|| | ||:|| ||:||| | ||  ||  :
Db       1567 LLLYNGQKTDGKGDFVSLALHNRHLEFCYDLGKGAAVIRSKEPIALGTWVRVFLERNGRK 1626

Qy        339 GSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGE 393
              |:|:|  |  ||  || ||:      |||   ||:||      |:    :|: | | :
Db       1627 GALQVGDGPRVLGESP-KSRKVPHTMLNLKEPLYIGGAPDFSKLARGAAVSSGFSGVIQL 1685

Qy        394 VSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDL 453
              ||: | :|      |  : : :  | ||    ||
Db       1686 VSLRGHQL------LTQEHVLRAVDVSPFADHPCTQA----------------------- 1716

Qy        454 CEHEENPCQLREPCLHGGTC---QGT-RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 509
                       |  |||:||:|   : |  |||  ||||  |::|           :| | |
Db       1717 ---------LGNPCLNGGSCVPREATYECLCPGGFSGLHCEKGL----------VEKSVG 1757

Qy        510 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASGLLLWQG 561
                          |   ||| |  :   |  | |:         || :||    ||:||
Db       1758 ------------DLETLAFDGRTYIEYLNAVIESEKALQSNHFELSLRTEATQGLVLW-- 1803

Qy        562 VEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSI 621
                :|:| :  |:::| : ||||   | |||    | |   :|   | |: | || | ||:
Db       1804 --IGKAAERADYMALAIVDGHLQLSYDLGSQPVVLRSTVKVNTNRWLRIRAHREHREGSL 1861

Qy        622 QVDGEELVSGRSPGPNVAVNAKGSVYIGGAP--DVATLTGGRFSSGITGCVKNLVLHSAR 679
              ||  |  |:| ||     ::  |::::||     |       : :|  ||::::|:
Db       1862 QVGNEAPVTGSSPLGATQLDTDGALWLGGLQKLPVGQALPKAYGTGFVGCLRDVVVGH-- 1919

Qy        680 PGAPPPQPLDLQHRAQAGANTRPCPS 705
                    : | |   |      ||||:
Db       1920 ------RQLHLLEDAVTKPELRPCPT 1939

RESULT 3
US-08-644-271-32
; Sequence 32, Application US/08644271
; Patent No. 5814478
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela, et al.
;    TITLE OF INVENTION:  NOVEL TYROSINE KINASE RECEPTORS
;    TITLE OF INVENTION:  AND LIGANDS
;    NUMBER OF SEQUENCES:  32
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/644,271
;      FILING DATE:  10-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 60/008,657
;      FILING DATE:  15-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 195A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  914-345-7400
;      TELEFAX:  914-345-7721
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  492 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
;    FEATURE:
;      NAME/KEY:  Human Agrin
;      LOCATION:  1...492
;      OTHER INFORMATION:
US-08-644-271-32

  Query Match             16.7%;  Score 640.5;  DB 2;  Length 492;
  Best Local Similarity   33.0%;  Pred. No. 6.6e-46;
  Matches  183;  Conservative   59;  Mismatches  173;  Indels  139;  Gaps   18;

Qy        202 CHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNT 261
              | |  |   | |   |:| |  | | ||| |  |:     |     |:| :||      :
Db         27 CQPNPCHGAAPCRVLPEG-GAQCECPLGREGTFCQ-----TASGQDGSGPFLA-DFNGFS 79

Qy        262 HHELR--------------LDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRY 307
              | |||              |:| |    | |:||::| |:    ||||||:    |||||
Db         80 HLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFRY 139

Qy        308 ELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQG-----LNL 362
              :|| | ||:|| ||: || | ||| ||  : |:|||  |  ||  || ||:      |||
Db        140 DLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESP-KSRKVPHTVLNL 198

Qy        363 HTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL----------DLTYSFLGSQG 412
                 ||:||      |: |  :|: | | :  ||: |::|          |:| || |
Db        199 KEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG--- 254

Qy        413 IGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLH 469
                      || |    || :||:|:|                            ||
Db        255 -------HPCTRASGHPCLNGASCVP---------------------------REAAY- 278

Qy        470 GGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFP 529
                      |||  ||||| |::|           :| | |            |   |||
Db        279 -------VCLCPGGFSGPHCEKGL----------VEKSAG------------DVDTLAFD 309

Qy        530 GHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQGKDF 573
              |  |   |  | |:                 || :||    ||:||     |:| :  |:
Db        310 GRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATERADY 365
```

```
Qy        574 ISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRS 633
              ::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:| |
Db        366 VALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSS 425

Qy        634 PGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQ 691
              |     ::  |::::|| |:  |       : :|  ||::::|:           || |
Db        426 PLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLL 477

Qy        692 HRAQAGANTRPCPS 705
                |      ||||:
Db        478 EDAVTKPELRPCPT 491

RESULT 4
US-09-077-955-36
; Sequence 36, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;   LENGTH: 492
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-36

  Query Match             16.7%;  Score 640.5;  DB 4;  Length 492;
  Best Local Similarity   33.0%;  Pred. No. 6.6e-46;
  Matches  183;  Conservative   59;  Mismatches  173;  Indels  139;  Gaps   18;

Qy        202 CHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNT 261
              | |  |   | |   |:| |  | | ||| |  |:     |     |:| :||      :
Db         27 CQPNPCHGAAPCRVLPEG-GAQCECPLGREGTFCQ-----TASGQDGSGPFLA-DFNGFS 79

Qy        262 HHELR--------------LDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRY 307
              | |||              |:| |    | |:||::| |:    ||||||:    |||||
Db         80 HLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFRY 139

Qy        308 ELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQG-----LNL 362
              :|| | ||:|| ||: || | ||| ||  : |:|||  |  ||  || ||:      |||
Db        140 DLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESP-KSRKVPHTVLNL 198

Qy        363 HTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL----------DLTYSFLGSQG 412
                 ||:||      |: |  :|: | | :  ||: |::|          |:| || |
Db        199 KEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG--- 254

Qy        413 IGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLH 469
                      || |    || :||:|:|                            ||
Db        255 -------HPCTRASGHPCLNGASCVP---------------------------REAAY- 278

Qy        470 GGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFP 529
                      |||  ||||| |::|           :| | |            |   |||
Db        279 -------VCLCPGGFSGPHCEKGL----------VEKSAG------------DVDTLAFD 309

Qy        530 GHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQGKDF 573
              |  |   |  | |:                 || :||    ||:||     |:| :  |:
Db        310 GRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATERADY 365

Qy        574 ISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRS 633
              ::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:| |
Db        366 VALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSS 425
```

```
Qy        634 PGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQ 691
              |     ::  |::::|| |:  |       : :|  ||::::|:           || |
Db        426 PLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLL 477

Qy        692 HRAQAGANTRPCPS 705
                |      ||||:
Db        478 EDAVTKPELRPCPT 491

RESULT 5
US-09-077-955-25
; Sequence 25, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 456
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-25

  Query Match             15.2%;  Score 579.5;  DB 4;  Length 456;
  Best Local Similarity   32.7%;  Pred. No. 8.8e-41;
  Matches  168;  Conservative   57;  Mismatches  156;  Indels  133;  Gaps   16;

Qy        242 TTPSLSGAGSYLALPALTNTHHELR--------------LDVEFKPLAPDGVLLFSGGKS 287
              |     |:| :||      :| |||              |:| |    | |:||::| |:
Db         25 TASGQDGSGPFLA-DFNGFSHLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKT 83

Qy        288 GPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGR 347
                  ||||||:    |||||:|| | ||:|| ||: || | ||| ||  : |:|||  |
Db         84 DGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGP 143

Qy        348 PVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL- 401
               ||  || ||:      |||   ||:||      |: |  :|: | | :  ||: |::|
Db        144 RVLGESP-KSRKVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLL 202

Qy        402 ---------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGF 449
                       |:| || |           || |    || :||:|:|
Db        203 TPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP-------------- 237

Qy        450 KGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGG 509
                            ||            |||  ||||| |::|           :| | |
Db        238 --------------REAAY--------VCLCPGGFSGPHCEKGL----------VEKSAG 265

Qy        510 NDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVRTSTA 553
                          |   ||| |  |   |  | |:                 || :||
Db        266 ------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEAT 313

Qy        554 SGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTAL 613
               ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | || |
Db        314 QGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAH 369

Qy        614 REGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVK 671
              || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||::
Db        370 REQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLR 429

Qy        672 NLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              ::|:           || |   |      ||||:
Db        430 DVVVGR--------HPLHLLEDAVTKPELRPCPT 455
```

```
RESULT 6
US-09-077-955-26
; Sequence 26, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 440
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-26

  Query Match             15.1%;  Score 576;  DB 4;  Length 440;
  Best Local Similarity   32.9%;  Pred. No. 1.6e-40;
  Matches  163;  Conservative   55;  Mismatches  145;  Indels  132;  Gaps   15;

Qy        261 THHELR--------------LDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFR 306
              :| |||              |:| |    | |:||::| |:    ||||||:    ||||
Db         27 SHLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFR 86

Qy        307 YELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQG-----LN 361
              |:|| | ||:|| ||: || | ||| ||  : |:|||  |  ||  || ||:      ||
Db         87 YDLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESP-KSRKVPHTVLN 145

Qy        362 LHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL----------DLTYSFLGSQ 411
              |   ||:||      |: |  :|: | | :  ||: |::|          |:| || |
Db        146 LKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG-- 202

Qy        412 GIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCL 468
                       || |    || :||:|:|                            ||
Db        203 --------HPCTRASGHPCLNGASCVP----------------------------REAAY 226

Qy        469 HGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAF 528
                       |||  ||||| |::|           :| | |            |   |||
Db        227 --------VCLCPGGFSGPHCEKGL----------VEKSAG------------DVDTLAF 256

Qy        529 PGHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQGKD 572
               |  |   |  | |:                 || :||    ||:||     |:| :  |
Db        257 DGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATERAD 312

Qy        573 FISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGR 632
              :::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:|
Db        313 YVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGS 372

Qy        633 SPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 690
              ||     ::  |::::|| |:  |       : :|  ||::::|:           || |
Db        373 SPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHL 424

Qy        691 QHRAQAGANTRPCPS 705
                 |      ||||:
Db        425 LEDAVTKPELRPCPT 439

RESULT 7
US-09-077-955-27
; Sequence 27, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
```

```
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
;   LENGTH: 390
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-27

  Query Match             13.8%;  Score 528;  DB 4;  Length 390;
  Best Local Similarity   33.0%;  Pred. No. 1.6e-36;
  Matches  149;  Conservative   50;  Mismatches  134;  Indels  118;  Gaps   14;

Qy        291 EDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVL 350
              :| ||||:    |||||:|| | ||:|| ||: || | ||| ||  : |:|||  |  ||
Db         21 DDKVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVL 80

Qy        351 RSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL---- 401
                || ||:      |||   ||:||      |: |  :|: | | :  ||: |::|
Db         81 GESP-KSRKVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPE 139

Qy        402 ------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGD 452
                    |:| || |           || |    || :||:|:|
Db        140 HVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP----------------- 171

Qy        453 LCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDA 512
                         ||            |||  ||||| |::|           :| | |
Db        172 -----------REAAY--------VCLCPGGFSGPHCEKGL----------VEKSAG--- 199

Qy        513 PGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEVRTSTASGL 556
                       |   ||| |  |   |  | |:                 || :||    ||
Db        200 ---------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGL 250

Qy        557 LLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREG 616
              :||     |:| :  |:::| : ||||   | |||    | |  |:|   | || | ||
Db        251 VLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQ 306

Qy        617 RRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLV 674
              | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||::::|
Db        307 REGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVV 366

Qy        675 LHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              :           || |   |      ||||:
Db        367 VGR--------HPLHLLEDAVTKPELRPCPT 389

RESULT 8
US-09-188-930-183
; Sequence 183, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 183
;   LENGTH: 771
;   TYPE: PRT
;   ORGANISM: Rat
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (717)...(717)
US-09-188-930-183

  Query Match             10.4%;  Score 397.5;  DB 3;  Length 771;
  Best Local Similarity   27.6%;  Pred. No. 5e-25;
  Matches  145;  Conservative   64;  Mismatches  185;  Indels  131;  Gaps   28;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |||| |::|  || |  :       |    |
Db        315 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCVCPEGYSGLFCEFSPPMVFLRTSPCDNFDC 373

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   || |      |:|  |  | :||: |:|:   |    ||| :|:      |
Db        374 QNGAQCIIRVNEP-----ICQCLPGYLGEKCEKLVSVSI--LVNKESYLQIPSAKVRPQT 426

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | :::  : | :   |: ||  |  : |
Db        427 NITLQIATDED------SGILLYKGDK-----DHIAVESIEG-IRASYDTGSHPASAIYS 474

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |:   || |:|| | : ::  |   ||  : ||:||:
Db        475 VETINDGNFHIVELLTLDSSLSLSVDGGSPKIITNLSKQSTLNFDSPLYVGGMPGKNNVA 534

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : :  :  |    || :: | || ||
Db        535 SLRQAPGQN-GTSFHGCIRNLYINSELQDFRKVPMQTGILPGC---EPCHKKVCAHG-TC 589

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | | :|: | ||:   | || |   |:| |||         | || |  |
Db        590 QPSSQSGFTCECEEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 647

Qy        488 RCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSRSLPEVPETIEL 546
               | :                              :| |   ||  | :|
Db        648 LCDE-----------------------------EEDLFNPLPGDQVQAR----------- 667

Qy        547 EVRTSTASGLLLW--QGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPIND 604
              ||:   |   |||  | :  |:  |      :  :   :   ||  |   | :|    ||
Db        668 EVQALWARAALLWMQQWIHRGQLTQ-----RISCRGERIRDYYQ--SSRVRCLS----ND 716

Qy        605 GEWHRVTA------------LREGRR--------GSIQVDGEELV 629
               |  |:              ||  ||        ||  ||  | |
Db        717 XEVSRLECRGGCAGGQCCGPLRSKRRKYSFECTDGSSFVDEVEKV 761

RESULT 9
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
;  APPLICANT: Goodman, Corey
;  APPLICANT: Kid, Thomas
;  APPLICANT: Brose, Katja
;  APPLICANT: Tessier-Lavigne, Marc
;  TITLE OF INVENTION: Modulating Robo: Ligand Interactions
;  FILE REFERENCE: B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/191,647
;  CURRENT FILING DATE: 1998-11-13
;  EARLIER APPLICATION NUMBER: 60/065,544
;  EARLIER FILING DATE: 1997-11-14
;  EARLIER APPLICATION NUMBER: 60/081,057
;  EARLIER FILING DATE: 1998-04-07
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-191-647-2
```

```
  Query Match             10.4%;  Score 397.5;  DB 3;  Length 1525;
  Best Local Similarity   29.7%;  Pred. No. 1.3e-24;
  Matches  108;  Conservative   58;  Mismatches  141;  Indels   57;  Gaps   18;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |:|| |::|  || |  :       |    |
Db       1068 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDC 1126

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   :| |      |:|  |  | :||: |:|    :    ||| :|:      |
Db       1127 QNGAQCIVRINEP-----ICQCLPGYQGEKCEKLVSV---NFINKESYLQIPSAKVRPQT 1178

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | ::: :  | :   |: ||  |  : |
Db       1179 NITLQIATDED------SGILLYKGDK-----DHIAVELYRGRVRASYDTGSHPASAIYS 1227

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |::   | |:|| | : ::  |   ||  : ||:||:
Db       1228 VETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVA 1287

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : :  :  |    || :: | || ||
Db       1288 SLRQAPGQN-GTSFHGCIRNLYINSELQDFQKVPMQTGILPGC---EPCHKKVCAHG-TC 1342

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | |::|: | ||:   | || |   |:| |||         | || |  |
Db       1343 QPSSQAGFTCECQEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 1400

Qy        488 RCQQ 491
               | :
Db       1401 LCDE 1404


RESULT 10
US-09-540-245A-2
; Sequence 2, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,245A
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  60/065,544
;  PRIOR FILING DATE:  1997-11-14
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-540-245A-2

  Query Match             10.4%;  Score 397.5;  DB 3;  Length 1525;
  Best Local Similarity   29.7%;  Pred. No. 1.3e-24;
  Matches  108;  Conservative   58;  Mismatches  141;  Indels   57;  Gaps   18;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |:|| |::|  || |  :       |    |
Db       1068 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDC 1126

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   :| |      |:|  |  | :||: |:|    :    ||| :|:      |
Db       1127 QNGAQCIVRINEP-----ICQCLPGYQGEKCEKLVSV---NFINKESYLQIPSAKVRPQT 1178

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | ::: :  | :   |: ||  |  : |
Db       1179 NITLQIATDED------SGILLYKGDK-----DHIAVELYRGRVRASYDTGSHPASAIYS 1227

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |::   | |:|| | : ::  |   ||  : ||:||:
Db       1228 VETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVA 1287

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : :  :  |    || :: | || ||
Db       1288 SLRQAPGQN-GTSFHGCIRNLYINSELQDFQKVPMQTGILPGC---EPCHKKVCAHG-TC 1342

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | |::|: | ||:   | || |   |:| |||         | || |  |
Db       1343 QPSSQAGFTCECQEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 1400

Qy        488 RCQQ 491
               | :
Db       1401 LCDE 1404


RESULT 11
US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,153
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  09/191,647
;  PRIOR FILING DATE:  1998-11-13
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-540-153-2

  Query Match             10.4%;  Score 397.5;  DB 3;  Length 1525;
  Best Local Similarity   29.7%;  Pred. No. 1.3e-24;
  Matches  108;  Conservative   58;  Mismatches  141;  Indels   57;  Gaps   18;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |:|| |::|  || |  :       |    |
Db       1068 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDC 1126

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   :| |      |:|  |  | :||: |:|    :    ||| :|:      |
Db       1127 QNGAQCIVRINEP-----ICQCLPGYQGEKCEKLVSV---NFINKESYLQIPSAKVRPQT 1178

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | ::: :  | :   |: ||  |  : |
Db       1179 NITLQIATDED------SGILLYKGDK-----DHIAVELYRGRVRASYDTGSHPASAIYS 1227

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |::   | |:|| | : ::  |   ||  : ||:||:
Db       1228 VETINDGNFHIVELLALDQSLSLSVDGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVA 1287

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : :  :  |    || :: | || ||
Db       1288 SLRQAPGQN-GTSFHGCIRNLYINSELQDFQKVPMQTGILPGC---EPCHKKVCAHG-TC 1342

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | |::|: | ||:   | || |   |:| |||         | || |  |
Db       1343 QPSSQAGFTCECQEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 1400

Qy        488 RCQQ 491
               | :
```

```
Db       1401 LCDE 1404

RESULT 12
US-09-312-283C-396
; Sequence 396, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 396
;   LENGTH: 1529
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-312-283C-396

  Query Match             10.3%;  Score 394.5;  DB 4;  Length 1529;
  Best Local Similarity   30.2%;  Pred. No. 2.4e-24;
  Matches  110;  Conservative   54;  Mismatches  143;  Indels   57;  Gaps   18;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |||| |::|  || |  :       |    |
Db       1072 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCVCPEGYSGLFCEFSPPMVLLRTSPCDNFDC 1130

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   || |      |:|  |  | :||: |:|   :     ||| :|:      |
Db       1131 QNGAQCIIRVNEP-----ICQCLPGYLGEKCEKLVSV---NFVNKESYLQIPSAKVRPQT 1182

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | ::: :  | :   |: ||  |  : |
Db       1183 NITLQIATDED------SGILLYKGDK-----DHIAVELYRGRVRASYDTGSHPASAIYS 1231

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |:   || |:|| | : ::  |   ||  : ||:||:
Db       1232 VETINDGNFHIVELLTLDSSLSLSVDGGSPKIITNLSKQSTLNFDSPLYVGGMPGKNNVA 1291

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : :  :  |    || :: | || ||
Db       1292 SLRQAPGQN-GTSFHGCIRNLYINSELQDFRKVPMQTGILPGC---EPCHKKVCAHG-TC 1346

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | | :|: | ||:   | || |   |:| |||         | || |  |
Db       1347 QPSSQSGFTCECEEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 1404

Qy        488 RCQQ 491
               | :
Db       1405 LCDE 1408

RESULT 13
US-09-312-283C-183
; Sequence 183, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 183
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-312-283C-183

  Query Match             10.3%;  Score 394;  DB 4;  Length 716;
  Best Local Similarity   28.3%;  Pred. No. 8.8e-25;
  Matches  126;  Conservative   59;  Mismatches  160;  Indels  100;  Gaps   23;

Qy        158 HC----PTCRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQAL------HCHPEAC 207
              ||      |:|  |:||  | |:  : | |||| |::|  || |  :       |    |
Db        315 HCDIDFDDCQDNKCKNGAHCTDA-VNGYTCVCPEGYSGLFCEFSPPMVFLRTSPCDNFDC 373

Qy        208 GPDATC---VNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPAL-----T 259
                 | |   || |      |:|  |  | :||: |:|    |    ||| :|:      |
Db        374 QNGAQCIIRVNEP-----ICQCLPGYLGEKCEKLVSVSI--LVNKESYLQIPSAKVRPQT 426

Qy        260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
              |   ::  | :       |:||: | |     | :::  : | :   |: ||  |  : |
Db        427 NITLQIATDED------SGILLYKGDK-----DHIAVESIEG-IRASYDTGSHPASAIYS 474

Qy        319 AEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE-----P 373
               | :  | :| |    |:   || |:|| | : ::  |   ||  : ||:||:
Db        475 VETINDGNFHIVELLTLDSSLSLSVDGGSPKIITNLSKQSTLNFDSPLYVGGMPGKNNVA 534

Qy        374 SVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATC 433
              |:  :|  |    | ||:  : :| :  |     : :  :  |    || :: | || ||
Db        535 SLRQAPGQN-GTSFHGCIRNLYINSELQDFRKVPMQTGILPGC---EPCHKKVCAHG-TC 589

Qy        434 MPAGEYEFQCLCRDGFKGDLCEHEEN-PCQLREPCLHGGTCQ-----GTRCLCLPGFSGP 487
               |: :  | | | :|: | ||:   | || |   |:| |||         | || |  |
Db        590 QPSSQSGFTCECEEGWMGPLCDQRTNDPC-LGNKCVH-GTCLPINAFSYSCKCLEGHGGV 647

Qy        488 RCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSRSLPEVPETIEL 546
               | :                              :| |   ||  | :|
Db        648 LCDE-----------------------------EEDLFNPLPGDQVQAR----------- 667

Qy        547 EVRTSTASGLLLW--QGVEVGEAGQ 569
              ||:   |   |||  | :  |:  |
Db        668 EVQALWARAALLWMQQWIHRGQLTQ 692

RESULT 14
US-09-182-024A-2
; Sequence 2, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
;  APPLICANT:  Connolly, Timothy
;  APPLICANT:  Rajput, Bhanu
;  TITLE OF INVENTION:  Human Slit Polypeptide and Polynucleotides Encoding
;  TITLE OF INVENTION:  Same
;  FILE REFERENCE:  640100-271
;  CURRENT APPLICATION NUMBER: US/09/182,024A
;  CURRENT FILING DATE:  1998-10-29
;  PRIOR APPLICATION NUMBER:  60/063,946
;  PRIOR FILING DATE:  1997-10-31
;  PRIOR APPLICATION NUMBER:  60/096,420
;  PRIOR FILING DATE:  1998-08-13
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-182-024A-2

  Query Match             10.2%;  Score 391.5;  DB 4;  Length 1523;
```

```
  Best Local Similarity   29.0%;  Pred. No. 4.3e-24;
  Matches  129;  Conservative   53;  Mismatches  158;  Indels  105;  Gaps   24;

Qy        133 CVRELRIQGE---EIVFH---DLNLTAH---------GIS----------HCPT----CR 163
              |:      ||   |:: |   :|||  |         | |           | |    |
Db       1020 CICPPNYTGELCDEVIDHCVPELNLCQHEAKCIPLDKGFSCECVPGYSGKLCETDNDDCV 1079

Qy        164 DRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHS------QALHCHPEACGPDATCV--- 214
                 |::| || |:  : | | || ||:|  |||       |   |    |   | |:
Db       1080 AHKCRHGAQCVDT-INGYTCTCPQGFSGPFCEHPPPMVLLQTSPCDQYECQNGAQCIVVQ 1138

Qy        215 NRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPL 274
                |     ||||  | :| |||: :||   :  |  ||: | |      :  : ::
Db       1139 QEP-----TCRCPPGFAGPRCEKLITV---NFVGKDSYVEL-ASAKVRPQANISLQVATD 1189

Qy        275 APDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYE-LGSGLAVLRSAEPLALGRWHRVSAE 333
                :|:||: |       |  ::| :  ||:   |: : |    : | | :  |::| |
Db       1190 KDNGILLYKGD-----NDPLALELYQGHVRLVYDSVSSPPTTVYSVETVNDGQFHSVEVV 1244

Qy        334 RLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMS----AHFRG 389
               ||:  :| |: | |       |   : ::: |||||:  |  ||     :      | |
Db       1245 TLNQTLNLVVDKGTPKSLGKFQKQPAVGINSPLYLGGIPTSTGLSALRQGTDRPLGGFHG 1304

Qy        390 CVGEVSVNGKRLD---LTYSFLG-SQGIGQC------------YDSSPCERQP------- 426
              |: || :| :  |   |    || | |   |             ||  || :|
Db       1305 CIHEVRINNELQDFKALPPQSLGVSPGCKSCTVCKHGLCRSVEKDSVVCECRPGWTGPLC 1364

Qy        427 ------------CQHGATCMPAGEYEFQCLCRDGFKGDLCEHEE---NPCQLREPCLHGG 471
                          | ||  |:  |  :  | | :|: ||||:::    | |   :   | |
Db       1365 DQEARDPCLGHRCHHG-KCVATGT-SYMCKCAEGYGGDLCDNKNDSANACSAFK--CHHG 1420

Qy        472 TC----QG-TRCLCLPGFSGPRCQQ 491
               |    ||   |||  |||||  |||
Db       1421 QCHISDQGEPYCLCQPGFSGEHCQQ 1445

RESULT 15
US-09-077-955-28
; Sequence 28, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
;  APPLICANT: Valenzuela et al., David M.
;  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
;  FILE REFERENCE: REG195-B-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/077,955A
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PCT/US96/20696
;  EARLIER FILING DATE: 1996-12-13
;  EARLIER APPLICATION NUMBER: 08/644,271
;  EARLIER FILING DATE: 1996-05-10
;  EARLIER APPLICATION NUMBER: 60/008,657
;  EARLIER FILING DATE: 1995-12-15
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-077-955-28

  Query Match             10.0%;  Score 382;  DB 4;  Length 338;
  Best Local Similarity   29.6%;  Pred. No. 3.1e-24;
  Matches  118;  Conservative   42;  Mismatches  125;  Indels  114;  Gaps   13;

Qy        338 DGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVN 397
              |   || |  |  |  |     |||   ||:||      |: |  :|: | | :  ||:
Db         22 DKGPRVLGESPKSRKVP--HTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLG 79

Qy        398 GKRL----------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCL 444
              |::|          |:| || |           || |    ||  :||:|:|
Db         80 GRQLLTPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP--------- 119

Qy        445 CRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHL 504
                                 ||            |||  ||||| |::|           :
Db        120 -------------------REAAY--------VCLCPGGFSGPHCEKGL----------V 142

Qy        505 EGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET----------------IELEV 548
              | | |            |   ||| |  |   |  | |:                 || :
Db        143 EKSAG------------DVDTLAFDGRTFVEYLNAVTESELANEIPVEKALQSNHFELSL 190

Qy        549 RTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWH 608
              ||    ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   |
Db        191 RTEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWL 246

Qy        609 RVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGI 666
              || | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|
Db        247 RVVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGF 306

Qy        667 TGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
               ||::::|:           || |   |      ||||:
Db        307 VGCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 337

Search completed: March  9, 2004, 17:23:32
Job time : 29.6355 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:11:22 ; Search time 85.8465 Seconds
                                           (without alignments)
                                           2320.373 Million cell updates/sec

Title:          US-10-006-011A-3
Perfect score:  3825
Sequence:       1 EIKITFRPDSADGMLLYNGQ..........QPLDLQHRAQAGANTRPCPS 705

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
               1:  geneseqp1980s:*
               2:  geneseqp1990s:*
               3:  geneseqp2000s:*
               4:  geneseqp2001s:*
               5:  geneseqp2002s:*
               6:  geneseqp2003as:*
               7:  geneseqp2003bs:*
               8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result              Query
  No.    Score      Match Length DB ID                          Description
------------------------------------------------------------------------------
   1     3825   100.0   4391  6  AAE34390                 Aae34390 Human per
   2     3815    99.7   4393  4  AAB31889                 Aab31889 Amino aci
   3     3795    99.2   4436  4  ABG23265                 Abg23265 Novel hum
   4     1030    26.9    195  4  AAB31890                 Aab31890 Amino aci
   5    901.5    23.6   1931  6  ABU52400                 Abu52400 Human GPC
   6    891.5    23.3   2053  7  ADC39154                 Adc39154 Novel hum
   7    849.5    22.2   2143  7  ADC39164                 Adc39164 Novel hum
   8      840    22.0    671  6  ABP58231                 Abp58231 Human cel
   9      840    22.0   1009  7  ADE28105                 Ade28105 Human NTR
  10      832    21.8    819  5  ABB72291                 Abb72291 Rat prote
  11    812.5    21.2   4072  4  ABB63614                 Abb63614 Drosophil
  12    664.5    17.4   1298  4  ABB61948                 Abb61948 Drosophil
  13    640.5    16.7    492  2  AAW26609                 Aaw26609 Human agr
  14    635.5    16.6   1566  7  ADC39166                 Adc39166 Novel hum
  15      628    16.4    463  7  ADB64904                 Adb64904 Human pro
  16    625.5    16.4   1565  7  ADC39156                 Adc39156 Novel hum
  17      605    15.8    432  4  AAU18102                 Aau18102 Novel hum
  18      605    15.8    432  4  AAU16938                 Aau16938 Human nov
  19      605    15.8    432  4  ABB10233                 Abb10233 Human cDN
  20      605    15.8    432  4  AAU19905                 Aau19905 Novel hum
  21      605    15.8    432  5  ABJ05729                 Abj05729 Novel hum
  22      605    15.8    432  5  ABP66820                 Abp66820 Human pol
  23    587.5    15.4    416  2  AAY73993                 Aay73993 Human pro
  24    576.5    15.1   1039  4  ABB63124                 Abb63124 Drosophil
  25      556    14.5   1741  5  ABP43859                 Abp43859 Human mRN
```

```
  26    543.5    14.2    375  4  AAU07421                 Aau07421 Novel hum
  27      441    11.5   3313  7  ADE54405                 Ade54405 Rat Prote
  28      441    11.5   3313  7  ADE54409                 Ade54409 Rat Prote
  29    420.5    11.0   1712  4  AAU28190                 Aau28190 Novel hum
  30    412.5    10.8   1327  4  AAM78871                 Aam78871 Human pro
  31      412    10.8   1522  4  AAM79855                 Aam79855 Human pro
  32      407    10.6   1508  2  AAY27141                 Aay27141 Human sli
  33      407    10.6   1508  2  AAW96706                 Aaw96706 Protein s
  34      407    10.6   1508  2  AAY04138                 Aay04138 Human sli
  35      407    10.6   1534  2  AAW46966                 Aaw46966 Amino aci
  36      407    10.6   1534  2  AAY27144                 Aay27144 Human sli
  37      407    10.6   1534  2  AAW96707                 Aaw96707 Protein s
  38      407    10.6   1534  2  AAY04139                 Aay04139 Human sli
  39      407    10.6   1534  7  ADE54473                 Ade54473 Human Pro
  40    401.5    10.5   1514  7  ADD49063                 Add49063 Human NOV
  41      398    10.4   1438  5  AAE17600                 Aae17600 Human ext
  42    397.5    10.4   1503  2  AAY27142                 Aay27142 Human sli
  43    397.5    10.4   1503  2  AAW96701                 Aaw96701 Slit-like
  44    397.5    10.4   1525  2  AAY17499                 Aay17499 Human Sli
  45    397.5    10.4   1525  7  ADB85322                 Adb85322 Human ner

                              ALIGNMENTS

RESULT 1
AAE34390
ID   AAE34390 standard; protein; 4391 AA.
XX
AC   AAE34390;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Human perlecan protein.
XX
KW   Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX
OS   Homo sapiens.
XX
PN   WO200295415-A2.
XX
PD   28-NOV-2002.
XX
PF   22-MAY-2002; 2002WO-EP005612.
XX
PR   23-MAY-2001; 2001GB-00012626.
XX
PA   (OSTE-) OSTEOMETER BIO TECH AS.
XX
PI   Christgau S,  Henriksen DB,  Cloos PAC;
XX
DR   WPI; 2003-140389/13.
XX
PT   An assay for the diagnosis or assessment of the severity of
PT   osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
PT   or optically inverted protein in a sample.
XX
PS   Disclosure; Page 46-67; 106pp; English.
XX
CC   The invention relates to an assay for the diagnosis or assessment of the
CC   severity of osteoarthritis or rheumatoid arthritis. The assay involves
CC   measuring (in a biological sample) the amount or presence of an
CC   isomerised or optically inverted protein or one or more isomerised or
CC   optically inverted fragments from proteins such as perlecan, biglycan,
CC   decorin, fibrillin-1 or protocadherin. The assay is useful for the
CC   diagnosis or assessment of the severity of osteoarthritis or rheumatoid
CC   arthritis. The present sequence is human perlecan protein
XX
SQ   Sequence 4391 AA;

  Query Match             100.0%;  Score 3825;  DB 6;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-233;
  Matches  705;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3687 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3746

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3747 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3806

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3807 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3866

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3867 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3926

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3927 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3987 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4046

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4047 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4106

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4107 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4166

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4167 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4226

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4227 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4286

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4287 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 4346

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4347 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
AAB31889
ID   AAB31889 standard; protein; 4393 AA.
XX
AC   AAB31889;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of a human protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 138-152; 209pp; French.
XX
CC   The present sequence represents a human protein, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 4393 AA;

  Query Match             99.7%;  Score 3815;  DB 4;  Length 4393;
  Best Local Similarity   99.6%;  Pred. No. 1.1e-232;
  Matches  702;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3689 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3748

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3749 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3808

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3809 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3868

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3869 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3928

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3929 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3988

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       3989 GHLEFRYELGSGLAVLRTAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4048

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4049 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4108

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       4109 PCERQPCQHGATCMPAGEYEFQCLCRDGIKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4168

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4169 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4228
```

```
Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4229 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 4288

Qy        601 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 660
              |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db       4289 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSIYIGGAPDVATLTGG 4348

Qy        661 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db       4349 RFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4393

RESULT 3
ABG23265
ID   ABG23265 standard; protein; 4436 AA.
XX
AC   ABG23265;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23256.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87452.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 53624; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4436 AA;

  Query Match             99.2%;  Score 3795;  DB 4;  Length 4436;
  Best Local Similarity   98.2%;  Pred. No. 2.1e-231;
  Matches  701;  Conservative    1;  Mismatches    2;  Indels   10;  Gaps    1;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3722 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 3781

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3782 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGA 3841

Qy        121 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3842 IPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSESSS 3901

Qy        181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3902 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3961

Qy        241 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3962 VTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 4021

Qy        301 GHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 360
              |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db       4022 GHLEFRYELGSGLAVLRTAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGL 4081

Qy        361 NLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 420
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       4082 NLHTLLYLGGVEPSVPLSPATNMGAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSS 4141

Qy        421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4142 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLC 4201

Qy        481 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4202 LPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEV 4261

Qy        541 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVF----------RYQLG 590
              |||||||||||||||||||||||||||||||||||||||||||||          |||||
Db       4262 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFSSYFLGLDDCRYQLG 4321

Qy        591 SGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGG 650
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       4322 SGEARLVSEDPFNDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGG 4381

Qy        651 APDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4382 APDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCP 4435

RESULT 4
AAB31890
ID   AAB31890 standard; protein; 195 AA.
XX
AC   AAB31890;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Amino acid sequence of the C-terminal of the human perlecan protein.
XX
KW   Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW   ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW   neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW   Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW   rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200105422-A2.
XX
PD   25-JAN-2001.
XX
PF   17-JUL-2000; 2000WO-FR002057.
XX
PR   15-JUL-1999;   99FR-00009372.
XX
PA   (INMR ) BIOMERIEUX STELHYS.
XX
PI   Roecklin D,  Kolbe H,  Charles M,  Malcus C,  Santoro L,  Perron H;
XX
DR   WPI; 2001-159475/16.
DR   N-PSDB; AAF54728.
XX
PT   Detecting, preventing and treating degenerative, neurological and
PT   autoimmune diseases, particularly multiple sclerosis, using specified
PT   polypeptides or related nucleic acid or ligand.
XX
PS   Claim 1; Page 152-153; 209pp; French.
XX
CC   The present sequence represents a human polypeptide, which is used in the
CC   method of the invention. The specification describes a method which uses
CC   at least one polypeptide or polynucleotide sequence belonging to the
CC   perlecan, precursor of the retinol-binding plasma protein, precursor of
CC   the ganglioside GM2 activator, calgranulin B or saposin B protein
CC   families. The method is used for detecting, preventing or treating a
CC   degenerative, neurological and/or auto-immune disease. The
CC   polynucleotides and polypeptides are used for diagnosis, prognosis,
CC   prevention and treatment of multiple sclerosis (in its various forms and
CC   phases). They may also be useful in cases of e.g. Alzheimer's and
CC   Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC   polyarthritis and lupus erythematosus, including use as vaccines and in
CC   gene therapy (expression of sense or antisense sequences). They can also
CC   be used to assess efficacy of potential therapeutic agents, particularly
CC   compounds that reduce or inhibit toxicity towards glial cells
XX
SQ   Sequence 195 AA;

  Query Match             26.9%;  Score 1030;  DB 4;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-57;
  Matches  195;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        511 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQG 60

Qy        571 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 KDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVS 120

Qy        631 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL 180

Qy        691 QHRAQAGANTRPCPS 705
              |||||||||||||||
Db        181 QHRAQAGANTRPCPS 195


RESULT 5
ABU52400
ID   ABU52400 standard; protein; 1931 AA.
XX
AC   ABU52400;
XX
DT   03-MAR-2003  (first entry)
XX
DE   Human GPCR related protein NOV40a.
XX
KW   Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytostatic.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200279398-A2.
XX
PD   10-OCT-2002.
XX
PF   08-MAR-2002; 2002WO-US007355.
XX
PR   08-MAR-2001; 2001US-0274194P.
PR   08-MAR-2001; 2001US-0274281P.
PR   08-MAR-2001; 2001US-0274322P.
PR   09-MAR-2001; 2001US-0274849P.
PR   13-MAR-2001; 2001US-0275578P.
PR   13-MAR-2001; 2001US-0275579P.
PR   13-MAR-2001; 2001US-0275601P.
PR   14-MAR-2001; 2001US-0276000P.
PR   16-MAR-2001; 2001US-0276776P.
PR   19-MAR-2001; 2001US-0276994P.
PR   20-MAR-2001; 2001US-0277239P.
PR   20-MAR-2001; 2001US-0277327P.
PR   20-MAR-2001; 2001US-0277338P.
PR   21-MAR-2001; 2001US-0277791P.
PR   22-MAR-2001; 2001US-0277833P.
PR   23-MAR-2001; 2001US-0278152P.
PR   26-MAR-2001; 2001US-0278894P.
PR   27-MAR-2001; 2001US-0278999P.
PR   27-MAR-2001; 2001US-0279036P.
PR   30-MAR-2001; 2001US-0280233P.
PR   02-APR-2001; 2001US-0280802P.
PR   02-MAY-2001; 2001US-0288052P.
PR   02-MAY-2001; 2001US-0288066P.
PR   02-MAY-2001; 2001US-0288228P.
PR   17-MAY-2001; 2001US-0291766P.
PR   07-JUN-2001; 2001US-0296693P.
PR   08-JUN-2001; 2001US-0296856P.
PR   05-JUL-2001; 2001US-0303230P.
PR   05-JUL-2001; 2001US-0303237P.
PR   08-AUG-2001; 2001US-0310913P.
PR   13-AUG-2001; 2001US-0311978P.
PR   14-AUG-2001; 2001US-0312191P.
PR   16-AUG-2001; 2001US-0312916P.
PR   17-AUG-2001; 2001US-0313182P.
PR   20-AUG-2001; 2001US-0313626P.
PR   21-AUG-2001; 2001US-0314018P.
PR   27-AUG-2001; 2001US-0315227P.
PR   10-SEP-2001; 2001US-0318403P.
PR   10-SEP-2001; 2001US-0318510P.
PR   14-SEP-2001; 2001US-0322296P.
PR   14-SEP-2001; 2001US-0322360P.
PR   27-SEP-2001; 2001US-0325378P.
PR   09-NOV-2001; 2001US-0332486P.
PR   09-NOV-2001; 2001US-0345399P.
PR   07-MAR-2002; 2002US-00094886.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Kekuda R,  Tchernev VT,  Liu X,  Spytek KA,  Patturajan M;
PI   Burgess CE,  Vernet CAM,  Li L,  Gorman L,  Malyankar UM,  Boldog FL;
PI   Guo X,  Shenoy SG,  Padigaru M,  Taupier RJ,  Miller CE,  Casman SJ;
PI   Pena CEA,  Gangolli EA,  Gusev V,  Smithson G,  Zerhusen BD,  Gerlach V;
PI   Pochart PF,  Fernandes ER,  Shimkets RA,  Rastelli L,  Spaderna SK;
PI   Larochelle WJ,  Zhong M,  Khramtsov NV,  Voss EZ,  Herrmann JL;
XX
DR   WPI; 2003-058423/05.
DR   N-PSDB; ABX70491.
XX
PT   NOVX polypeptides and polynucleotides, useful for treating a syndrome
PT   related to a human disease associated with the NOVX polypeptide e.g.,
PT   cancer.
XX
PS   Claim 1; Page 223; 413pp; English.
XX
CC   The present invention relates to the isolation of novel human
CC   polypeptides referred to as NOVX (NOV1-NOV44), variants of these
```

```
CC   proteins, and the polynucleotide sequences encoding them. The NOVX
CC   proteins of the invention are G-protein coupled receptor (GPCR) related
CC   proteins. The sequences of the invention are useful in the manufacture of
CC   a medicament for treating a syndrome related to a human disease
CC   associated with the polypeptides e.g. cancer. ABU52311-ABU52408 represent
CC   the human NOVX proteins of the invention
XX
SQ   Sequence 1931 AA;

  Query Match             23.6%;  Score 901.5;  DB 6;  Length 1931;
  Best Local Similarity   32.2%;  Pred. No. 2.5e-48;
  Matches  242;  Conservative   93;  Mismatches  266;  Indels  151;  Gaps   22;

Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              : : ||     |:|||||             |  ||::  |: || : ||| ||| | :
Db       1282 LALEFRALEPQGLLLYNGNA-----------RGKDFLALALLDGRVQLRFDTGSGPAVLT 1330

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD---Y 118
                 |:  | :| : | |   :|:| |    || | |     ||:|: :|::|| |:
Db       1331 SAVPVEPGQWHRLELSRHWRRGTLSVDGETPVLGESPSGTDGLNLDTDLFVGGVPEDQAA 1390

Qy        119 GAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNL------TAHGISHC--PTCRDRPCQNG 170
               |: :  : :|  ||:| | :  : :   :| :         |:  |    |   ||  |
Db       1391 VALERTFVGAGLRGCIRLLDVNNQRL---ELGIGPGAATRGSGVGECGDHPCLPNPCHGG 1447

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
                | : |:  : | || |  |  |   ::  | |  |   | |   |:| |  | | |||
Db       1448 APCQNLEAGRFHCQCPPGRVGPTCADEKS-PCQPNPCHGAAPCRVLPEG-GAQCECPLGR 1505

Qy        231 SGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELR--------------LDVEFKPLAP 276
               |  |:     |     |:| :||      :| |||              |:| |    |
Db       1506 EGTFCQ-----TASGQDGSGPFLA-DFNGFSHLELRGLHTFARDLGEKMALEVVFLARGP 1559

Qy        277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLN 336
               |:||::| |:    ||||||:    |||||:|| | ||:|| ||: || | ||| ||
Db       1560 SGLLLYNGQKTDGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNG 1619

Qy        337 KDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSV 396
              : |:|||  |  ||  ||     |||   ||:||      |: |  :|: | | :  ||:
Db       1620 RKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSL 1679

Qy        397 NGKRL----------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQC 443
               |::|          |:| || |           || |    || :||:|:|
Db       1680 GGRQLLTPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP-------- 1720

Qy        444 LCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWH 503
                                  ||            |||  ||||| |::|
Db       1721 --------------------REAAY--------VCLCPGGFSGPHCEKGL---------- 1742

Qy        504 LEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRTSTASG 555
              :| | |            |   ||| |  |   |  | |:         || :||    |
Db       1743 VEKSAG------------DVDTLAFDGRTFVEYLNAVTESEKALQSNHFELSLRTEATQG 1790

Qy        556 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALRE 615
              |:||     |:| :  |:::| : ||||   | |||    | |  |:|   | || | ||
Db       1791 LVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHRE 1846

Qy        616 GRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNL 673
               | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  ||::::
Db       1847 QREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDV 1906

Qy        674 VLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |:           || |   |      ||||:
Db       1907 VVGR--------HPLHLLEDAVTKPELRPCPT 1930


RESULT 6
ADC39154
ID   ADC39154 standard; protein; 2053 AA.
XX
AC   ADC39154;
XX
DT   18-DEC-2003  (first entry)
```

```
XX
DE   Novel human NOVX polypeptide  SEQ ID NO: 96.
XX
KW   antidiabetic; cytostatic; immunomodulator; anorectic; antilipemic;
KW   nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
KW   antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
KW   hemostatic; osteopathic; gene therapy.; NOVX; diabetes; obesity; cancer;
KW   lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
KW   wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
KW   cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
KW   hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
KW   Albright hereditary osteodystrophy.
XX
OS   Homo sapiens.
XX
PN   WO2003010327-A2.
XX
PD   06-FEB-2003.
XX
PF   02-MAY-2002; 2002WO-US014199.
XX
PR   02-MAY-2001; 2001US-0288063P.
PR   03-MAY-2001; 2001US-0288395P.
PR   07-MAY-2001; 2001US-0289087P.
PR   09-MAY-2001; 2001US-0289817P.
PR   09-MAY-2001; 2001US-0289818P.
PR   11-MAY-2001; 2001US-0290194P.
PR   14-MAY-2001; 2001US-0290753P.
PR   15-MAY-2001; 2001US-0291181P.
PR   16-MAY-2001; 2001US-0291243P.
PR   18-MAY-2001; 2001US-0292001P.
PR   21-MAY-2001; 2001US-0292374P.
PR   22-MAY-2001; 2001US-0292587P.
PR   23-MAY-2001; 2001US-0293107P.
PR   25-MAY-2001; 2001US-0293747P.
PR   29-MAY-2001; 2001US-0294109P.
PR   29-MAY-2001; 2001US-0294110P.
PR   30-MAY-2001; 2001US-0294434P.
PR   31-MAY-2001; 2001US-0294827P.
PR   12-JUL-2001; 2001US-0304879P.
PR   31-JUL-2001; 2001US-0308901P.
PR   14-AUG-2001; 2001US-0312270P.
PR   17-AUG-2001; 2001US-0313416P.
PR   10-SEP-2001; 2001US-0318463P.
PR   27-SEP-2001; 2001US-0325683P.
PR   18-OCT-2001; 2001US-0330292P.
PR   28-NOV-2001; 2001US-0333873P.
PR   03-DEC-2001; 2001US-0336909P.
PR   03-DEC-2001; 2001US-0337552P.
PR   21-FEB-2002; 2002US-0359245P.
PR   01-MAY-2002; 2002US-00136826.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Miller CE,   Kekuda R,   Malyankar UM,   Li L,   Pena CEA,   Spytek KA;
PI   Gorman L,   Guo X,   Fernandes ER,   Smithson G,   Stone DJ,   Zerhusen BD;
PI   Patturajan M,   Anderson DW,   Mezes PS,   Peyman JA,   Macdougall JR;
PI   Padigaru M,   Rastelli L,   Shenoy SG,   Gerlach VL,   Shimkets RA,   Zhong M;
PI   Edinger SR,   Ellerman K;
XX
DR   WPI; 2003-239445/23.
DR   N-PSDB; ADC39153.
XX
PT   New NOVX polypeptides and polynucleotides, useful in gene therapy,
PT   particularly for treating or preventing a syndrome associated with a
PT   human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
PT   hypertension or hemophilia.
XX
PS   Claim 1; SEQ ID NO 96; 748pp; English.
XX
CC   The invention relates to new isolated NOVX polypeptides, the genes
CC   encoding them or sequences having at least 95% identity to the amino acid
CC   or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
```

```
CC   particularly in the manufacture of a medicament for treating a syndrome
CC   associated with a human disease, which includes a pathology associated
CC   with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC   treating, preventing or alleviating pathology associated with NOVX
CC   polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC   polypeptide are especially useful for treating or preventing e.g.
CC   diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC   cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
CC   Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC   disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
CC   graft-versus-host disease or Albright hereditary osteodystrophy. The DNA
CC   encoding the protein is useful in gene therapy for treating the above
CC   conditions. These are also useful in developing powerful assay system for
CC   functional analysis of various human disorders, as well as in diagnostic
CC   applications. This sequence represents one of the NOVX proteins of the
CC   invention.
XX
SQ   Sequence 2053 AA;

  Query Match             23.3%;  Score 891.5;  DB 7;  Length 2053;
  Best Local Similarity   32.1%;  Pred. No. 1.2e-47;
  Matches  243;  Conservative   94;  Mismatches  263;  Indels  157;  Gaps   24;

Qy           2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
               : : ||     |:|||||             |  ||::  |: || : ||| ||| | :
Db        1400 LALEFRALEPQGLLLYNGNA-----------RGKDFLALALLDGRVQLRFDTGSGPAVLT 1448

Qy          62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD---Y 118
                  |:  | :| : | |   :|:| |    || | |     ||:|: :|::|| |:
Db        1449 SAVPVEPGQWHRLELSRHWRRGTLSVDGETPVLGESPSGTDGLNLDTDLFVGGVPEDQAA 1508

Qy         119 GAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNL------TAHGISHC--PTCRDRPCQNG 170
                |: :  : :|  ||:| | :  : :   :| :         |:  |    |   ||  |
Db        1509 VALERTFVGAGLRGCIRLLDVNNQRL---ELGIGPGAATRGSGVGECGDHPCLPNPCHGG 1565

Qy         171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
                 | : |:  : | || |  |  |   ::  | |  |   | |   |:| |  | | |||
Db        1566 APCQNLEAGRFHCQCPPGRVGPTCADEKS-PCQPNPCHGAAPCRVLPEG-GAQCECPLGR 1623

Qy         231 SGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELR--------------LDVEFKPLAP 276
                |  |:     |     |:| :||      :| |||              |:  |    |
Db        1624 EGTFCQ-----TASGQDGSGPFLA-DFNGFSHLELRGLHTIARDLGEKMALEAVFLARGP 1677

Qy         277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLN 336
                |:||::| |:    ||||||:    |||||:|| | ||:|| ||: || | ||| ||
Db        1678 SGLLLYNGQKTDGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNG 1737

Qy         337 KDGSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCV 391
               : |:|||  |  ||  || ||:      |||   ||:||      |: |  :|: | | :
Db        1738 RKGALRVGDGPRVLGESP-KSRKVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAI 1796

Qy         392 GEVSVNGKRL----------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGE 438
                 ||: |::|          |:| || |           || |    || :||:|:|
Db        1797 QLVSLGGRQLLTPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP--- 1842

Qy         439 YEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIA 498
                                        ||            |||  ||||| |::|
Db        1843 -------------------------REAAY--------VCLCPGGFSGPHCEKGL----- 1864

Qy         499 ESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRT 550
                    :| | |            |   ||| |  |   |  | |:         || :||
Db        1865 -----VEKSAG------------DVDTLAFDGRTFVEYLNAVTESEKALQSNHFELSLRT 1907

Qy         551 STASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRV 610
                   ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | ||
Db        1908 EATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRV 1963

Qy         611 TALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITG 668
                | || | ||:||  |  |:| ||     ::  |::::|| |:  |       : :|  |
Db        1964 VAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVG 2023

Qy         669 CVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
               |::::|:           || |   |      ||||:
```

```
Db        2024 CLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 2052

RESULT 7
ADC39164
ID   ADC39164 standard; protein; 2143 AA.
XX
AC   ADC39164;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Novel human NOVX polypeptide  SEQ ID NO: 106.
XX
KW   antidiabetic; cytostatic; immunomodulator; anorectic; antilipemic;
KW   nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
KW   antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
KW   hemostatic; osteopathic; gene therapy.; NOVX; diabetes; obesity; cancer;
KW   lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
KW   wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
KW   cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
KW   hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
KW   Albright hereditary osteodystrophy.
XX
OS   Homo sapiens.
XX
PN   WO2003010327-A2.
XX
PD   06-FEB-2003.
XX
PF   02-MAY-2002; 2002WO-US014199.
XX
PR   02-MAY-2001; 2001US-0288063P.
PR   03-MAY-2001; 2001US-0288395P.
PR   07-MAY-2001; 2001US-0289087P.
PR   09-MAY-2001; 2001US-0289817P.
PR   09-MAY-2001; 2001US-0289818P.
PR   11-MAY-2001; 2001US-0290194P.
PR   14-MAY-2001; 2001US-0290753P.
PR   15-MAY-2001; 2001US-0291181P.
PR   16-MAY-2001; 2001US-0291243P.
PR   18-MAY-2001; 2001US-0292001P.
PR   21-MAY-2001; 2001US-0292374P.
PR   22-MAY-2001; 2001US-0292587P.
PR   23-MAY-2001; 2001US-0293107P.
PR   25-MAY-2001; 2001US-0293747P.
PR   29-MAY-2001; 2001US-0294109P.
PR   29-MAY-2001; 2001US-0294110P.
PR   30-MAY-2001; 2001US-0294434P.
PR   31-MAY-2001; 2001US-0294827P.
PR   12-JUL-2001; 2001US-0304879P.
PR   31-JUL-2001; 2001US-0308901P.
PR   14-AUG-2001; 2001US-0312270P.
PR   17-AUG-2001; 2001US-0313416P.
PR   10-SEP-2001; 2001US-0318463P.
PR   27-SEP-2001; 2001US-0325683P.
PR   18-OCT-2001; 2001US-0330292P.
PR   28-NOV-2001; 2001US-0333873P.
PR   03-DEC-2001; 2001US-0336909P.
PR   03-DEC-2001; 2001US-0337552P.
PR   21-FEB-2002; 2002US-0359245P.
PR   01-MAY-2002; 2002US-00136826.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Miller CE,  Kekuda R,  Malyankar UM,  Li L,  Pena CEA,  Spytek KA;
PI   Gorman L,  Guo X,  Fernandes ER,  Smithson G,  Stone DJ,  Zerhusen BD;
PI   Patturajan M,  Anderson DW,  Mezes PS,  Peyman JA,  Macdougall JR;
PI   Padigaru M,  Rastelli L,  Shenoy SG,  Gerlach VL,  Shimkets RA,  Zhong M;
PI   Edinger SR,  Ellerman K;
XX
DR   WPI; 2003-239445/23.
DR   N-PSDB; ADC39163.
XX
```

```
PT   New NOVX polypeptides and polynucleotides, useful in gene therapy,
PT   particularly for treating or preventing a syndrome associated with a
PT   human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
PT   hypertension or hemophilia.
XX
PS   Claim 1; SEQ ID NO 106; 748pp; English.
XX
CC   The invention relates to new isolated NOVX polypeptides, the genes
CC   encoding them or sequences having at least 95% identity to the amino acid
CC   or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
CC   particularly in the manufacture of a medicament for treating a syndrome
CC   associated with a human disease, which includes a pathology associated
CC   with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC   treating, preventing or alleviating pathology associated with NOVX
CC   polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC   polypeptide are especially useful for treating or preventing e.g.
CC   diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC   cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
CC   Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC   disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
CC   graft-versus-host disease or Albright hereditary osteodystrophy. The DNA
CC   encoding the protein is useful in gene therapy for treating the above
CC   conditions. These are also useful in developing powerful assay system for
CC   functional analysis of various human disorders, as well as in diagnostic
CC   applications. This sequence represents one of the NOVX proteins of the
CC   invention.
XX
SQ   Sequence 2143 AA;

  Query Match             22.2%;  Score 849.5;  DB 7;  Length 2143;
  Best Local Similarity   29.0%;  Pred. No. 5.7e-45;
  Matches  246;  Conservative   95;  Mismatches  259;  Indels  247;  Gaps   26;

Qy          2 IKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIR 61
              : : ||     |:|||||             |  ||::  |: || : ||| ||| | :
Db       1400 LALEFRALEPQGLLLYNGNA-----------RGKDFLALALLDGRVQLRFDTGSGPAVLT 1448

Qy         62 HPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD---Y 118
                 |:  | :| : | |   :|:| |    || | |     ||:|: :|::|| |:
Db       1449 SAVPVEPGQWHRLELSRHWRRGTLSVDGETPVLGESPSGTDGLNLDTDLFVGGVPEDQAA 1508

Qy        119 GAIPKAGLSSGFIGCVRELRIQGEEIVFHDLNL------TAHGISHC--PTCRDRPCQNG 170
               |: :  : :|  ||:| | :  : :   :| :         |:  |    |   ||  |
Db       1509 VALERTFVGAGLRGCIRLLDVNNQRL---ELGIGPGAATRGSGVGECGDHPCLPNPCHGG 1565

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
                | : |:  : | || |  |  |   ::  | |  |   | |   |:| |  | | |||
Db       1566 APCQNLEAGRFHCQCPPGRVGPTCADEKS-PCQPNPCHGAAPCRVLPEG-GAQCECPLGR 1623

Qy        231 SGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELR--------------LDVEFKPLAP 276
               |  |:     |     |:| :||      :| |||              |:  |    |
Db       1624 EGTFCQ-----TASGQDGSGPFLA-DFNGFSHLELRGLHTIARDLGEKMALEAVFLARGP 1677

Qy        277 DGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLN 336
               |:||::| |:    ||||||:    |||||:|| | ||:|| ||: || | ||| ||
Db       1678 SGLLLYNGQKTDGKGDFVSLALRDRRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNG 1737

Qy        337 KDGSLRVNGGRPVLRSSPGKSQG-----LNLHTLLYLGGVEPSVPLSPATNMSAHFRGCV 391
              : |:|||  |  ||  || ||:      |||   ||:||      |: |  :|: | | :
Db       1738 RKGALRVGDGPRVLGESP-KSRKVPHTVLNLKEPLYVGGAPDFSKLARAAAVSSGFDGAI 1796

Qy        392 GEVSVNGKRL----------DLTYSFLGSQGIGQCYDSSPCER---QPCQHGATCMPAGE 438
                ||: |::|          |:| || |           || |    || :||:|:|
Db       1797 QLVSLGGRQLLTPEHVLRQVDVT-SFAG----------HPCTRASGHPCLNGASCVP--- 1842

Qy        439 YEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIA 498
                                       ||            |||  ||||| |::|
Db       1843 -------------------------REAAY--------VCLCPGGFSGPHCEKGL----- 1864

Qy        499 ESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPET--------IELEVRT 550
                   :| | |            |   ||| |  |   |  | |:         || :||
Db       1865 -----VEKSAG------------DVDTLAFDGRTFVEYLNAVTESEKALQSNHFELSLRT 1907
```

```
Qy        551 STASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRV 610
                  ||:||     |:| :  |:::| : ||||   | |||    | |  |:|   | ||
Db       1908 EATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRV 1963

Qy        611 TALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIG--------GAPDVATLTG--- 659
               | || | ||:||  |  |:| ||     ::  |::::|        | | |: : |
Db       1964 VAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGECFGETREGCPRVSDIRGTDS 2023

Qy        660 ---------------------------------GR------------------------- 661
                                               ||
Db       2024 TPQRPPLSQGACEPAGWALFSRCSPCSSQCCGAGRRVPRCGPPAGHLLVGVPISITESQP 2083

Qy        662 -----------------------FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGA 698
                                     : :|  ||::::|:           || |   |
Db       2084 GDSHCLCCRGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLLEDAVTKP 2135

Qy        699 NTRPCPS 705
                ||||:
Db       2136 ELRPCPT 2142


RESULT 8
ABP58231
ID   ABP58231 standard; protein; 671 AA.
XX
AC   ABP58231;
XX
DT   31-MAR-2003  (first entry)
XX
DE   Human cell adhesion and extracellular matrix protein 8.
XX
KW   Cell adhesion and extracellular matrix protein 8; CADECM-8; human;
KW   anti-HIV; virucide; antiallergic; antiinflammatory; antianaemic;
KW   antiparkinsonian; nootropic; anticonvulsant; antiinfertility;
KW   antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
KW   cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
KW   antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic;
KW   antiparasitic; anthelmintic; antipsoriatic; uropathic; ophthalmological;
KW   antirheumatic; haemostatic; antibacterial; protozoacide; fungicide;
KW   gynaecological; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200288322-A2.
XX
PD   07-NOV-2002.
XX
PF   01-MAY-2002; 2002WO-US013874.
XX
PR   02-MAY-2001; 2001US-0288290P.
PR   21-MAY-2001; 2001US-0292468P.
PR   15-JUN-2001; 2001US-0298616P.
PR   28-JUN-2001; 2001US-0301672P.
PR   04-JAN-2002; 2002US-0345008P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yue H,  Lee EA,  Duggan BM,  Thangavelu K,  Honchell CD,  Ding L;
PI   Hillman JL,  Baughn MR,  Kallick DA,  Lee S,  Warren BA,  Xu Y,  Tran UK;
PI   Lal PG,  Thornton M,  Hafalia AJA,  Yao MG,  Nguyen DB,  Gandhi AR;
PI   Khan FA,  Walhia NK,  Griffin JA,  Chinn AM,  Elliott VS,  Ramkumar J;
PI   Arvizu CS,  Forsythe IJ;
XX
DR   WPI; 2003-167112/16.
DR   N-PSDB; ABZ24585.
XX
PT   New human cell adhesion and extracellular matrix proteins, useful for
PT   diagnosing, treating or preventing autoimmune or inflammatory disorder
PT   (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
PT   cancer or hepatitis.
XX
PS   Claim 1; Page 157-159; 178pp; English.
XX
```

```
CC   The present sequence is the protein sequence of Incyte polypeptide
CC   2736276CD1 denoted human cell adhesion and extracellular matrix protein 8
CC   (CADECM-8). The protein is encoded by a clone isolated from a soft tissue
CC   tumour cDNA library. The invention provides CADECM-1 to -11 polypeptides
CC   (see ABP58224-34) and polynucleotides (see ABZ24578-88), vectors, host
CC   cells, antibodies, agonists and antagonists. These are useful for
CC   diagnosing, treating or preventing disorders associated with aberrant
CC   expression of CADECM, particularly cell proliferative disorders (e.g.
CC   arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal
CC   nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary
CC   thromobocytopaenia or cancer), developmental disorders (e.g. renal
CC   tubular acidosis, anaemia or mental retardation), neurological disorders
CC   (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive
CC   disorders (e.g. infertility or a disruption in the menstrual cycle), or
CC   autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune
CC   thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,
CC   glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease,
CC   Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
CC   osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid
CC   arthritis, Sjogren's syndrome, uveitis), or viral, bacterial, fungal,
CC   parasitic, protozoal or helminthic infections
XX
SQ   Sequence 671 AA;

  Query Match             22.0%;  Score 840;  DB 6;  Length 671;
  Best Local Similarity   29.6%;  Pred. No. 6e-45;
  Matches  214;  Conservative  107;  Mismatches  257;  Indels  144;  Gaps   20;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              :| : || :: ||:||| |:           : : ||:|  ::    :|||: |:|:| |
Db         72 QITLEFRAEAEDGLLLYCGENE---------HGRGDFMSLAIIRRSLQFRFNCGTGVAII 122

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD-YG 119
                 | : || :||| | |    | | : :  || | |||::  :     ||||| |  |
Db        123 VSETKIKLGGWHTVMLYRDGLNGLLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYW 182

Qy        120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN-------LTAHGISHCPT--CRDRPCQNG 170
               :   | : || |||: | : |  |   |:        |:   :  | :  | :  | :|
Db        183 LVRATGTNRGFQGCVQSLAVNGRRI---DMRPWPLGKALSGADVGECSSGICDEASCIHG 239

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   :: ||:|:|| ||
Db        240 GTCTAIKADSYICLCPLGF----------------------------------------- 258

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
               |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db        259 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 318

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: : |||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db        319 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 373

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        374 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 433

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        434 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 478

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        479 ----LH---CQKAIIEAIEIPQFIG---------------------------RSYLTYDN 504

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  :|:   |||||:|          |||||||:||
Db        505 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 553

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        554 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 613

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
```

```
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        614 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 666

Qy        702 PC 703
               |
Db        667 TC 668

RESULT 9
ADE28105
ID   ADE28105 standard; protein; 1009 AA.
XX
AC   ADE28105;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human NTRAN protein - SEQ ID 10.
XX
KW   human; neurotransmission-associated protein; NTRAN; cytostatic;
KW   immunomodulator; immune disorder; cancer; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO2003051902-A1.
XX
PD   26-JUN-2003.
XX
PF   12-DEC-2002; 2002WO-US040059.
XX
PR   14-DEC-2001; 2001US-0340798P.
PR   18-MAR-2002; 2002US-0365645P.
PR   25-MAR-2002; 2002US-0367662P.
PR   10-MAY-2002; 2002US-0379887P.
PR   31-MAY-2002; 2002US-0384639P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Baughn MR,  Bhatia U,  Blake JJ,  Burrill JD,  Elliott VS;
PI   Emerling BM,  Forsythe IJ,  Gietzen KJ,  Gorvad AE,  Griffin JA;
PI   Hafalia AJA,  Ho A,  Jackson AA,  Jiang X,  Kable AE,  Kearney L;
PI   Khare R,  Lee EA,  Lee S,  Lu DAM,  Marquis JP,  Lehr-Mason PM;
PI   Ramkumar J,  Richardson TW,  Sprague WW,  Tran UK,  Chawla NK;
PI   Warren BA,  Yue H,  Zheng W;
XX
DR   WPI; 2003-514037/48.
DR   N-PSDB; ADE28127.
XX
PT   New human neurotransmission-associated proteins (NTRAN) polypeptide,
PT   useful for preparing a composition for treating a disease associated with
PT   decreased expression or overexpression of NTRAN e.g., cancer.
XX
PS   Claim 1; SEQ ID NO 10; 261pp; English.
XX
CC   The invention relates to a novel isolated human neurotransmission-
CC   associated proteins (NTRAN) polypeptide. The polypeptide of the invention
CC   demonstrates cytostatic and immunomodulator activities and may be useful
CC   for preparing a composition for diagnosing or treating a disease or
CC   condition associated with decreased expression or overexpression of
CC   functional NTRAN including immune disorders or cancer, as well as during
CC   gene therapy procedures. The current sequence is that of the human NTRAN
CC   protein of the invention.
XX
SQ   Sequence 1009 AA;

  Query Match             22.0%;  Score 840;  DB 7;  Length 1009;
  Best Local Similarity   29.6%;  Pred. No. 9.6e-45;
  Matches  214;  Conservative  107;  Mismatches  257;  Indels  144;  Gaps   20;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              :| : || :: ||:||| |:           : : ||:|  ::    :|||: |:|:| |
Db        410 QITLEFRAEAEDGLLLYCGENE---------HGRGDFMSLAIIRRSLQFRFNCGTGVAII 460

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD-YG 119
```

```
                 | : || :||| | |    | | : :  || | |||::  :     ||||| |  |
Db        461 VSETKIKLGGWHTVMLYRDGLNGLLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYW 520

Qy        120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN-------LTAHGISHCPT--CRDRPCQNG 170
               :   | : || |||: | : |  |   |:        |:   :  | :  | :  | :|
Db        521 LVRATGTNRGFQGCVQSLAVNGRRI---DMRPWPLGKALSGADVGECSSGICDEASCIHG 577

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   ::  ||:|:|| ||
Db        578 GTCTAIKADSYICLCPLGF----------------------------------------- 596

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
               |  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db        597 KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFMEFEITFRPDSGDGVLLYSYDT 656

Qy        284 GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRV 343
              | |     ||:|: : |||:|||:: |||  |||| :|| || || :   |  |:| |:|
Db        657 GSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDPLTLGNWHELRVSRTAKNGILQV 711

Qy        344 NGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDL 403
              :  : |   : |    :  :| :::|||     :   : :   | | : :: :| : : :
Db        712 DKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGVLKPFSGSIQKIILNDRTIHV 771

Qy        404 TYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQL 463
               : |  : |:     : || | || || :| |  |  : | |  ||:|
Db        772 KHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GYDCDCPLGFEG------------ 816

Qy        464 REPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFH 521
                  ||   ||      : :| | |                             |  | :
Db        817 ----LH---CQKAIIEAIEIPQFIG---------------------------RSYLTYDN 842

Qy        522 DDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDG 581
               |      |   |||       : :  :|:   |||||:|          |||||||:||
Db        843 PDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 891

Qy        582 HLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVN 641
               ||| | |||| | ::     ||| |||| |:|:|: | | ||     :|:|||    :|
Db        892 ALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLN 951

Qy        642 AKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 701
                |::|:||  ::|  |  ::  |: ||: :  | :          : |   |  | |
Db        952 INGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNIN 1004

Qy        702 PC 703
               |
Db       1005 TC 1006


RESULT 10
ABB72291
ID   ABB72291 standard; protein; 819 AA.
XX
AC   ABB72291;
XX
DT   04-APR-2002  (first entry)
XX
DE   Rat protein isolated from skin cells SEQ ID NO: 503.
XX
KW   Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW   developmental defect; inflammatory disease; dermatological; vulnerary;
KW   immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX
OS   Rattus sp.
XX
PN   WO200190357-A1.
XX
PD   29-NOV-2001.
XX
PF   24-MAY-2001; 2001WO-NZ000099.
XX
PR   24-MAY-2000; 2000US-0206650P.
PR   25-JUL-2000; 2000US-0221232P.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Watson JD,  Strachan L,  Sleeman M,  Onrust R,  Murison JG;
PI   Kumble KD;
XX
DR   WPI; 2002-122020/16.
XX
PT   New polynucleotides and polypeptides encoded by the polynucleotides
PT   isolated from skin cells, useful for treating skin wounds, cancers,
PT   growth and developmental defects, inflammatory diseases, or for
PT   modulating immune responses.
XX
PS   Claim 4; Page 308-310; 466pp; English.
XX
CC   The present invention provides the protein and coding sequences of cDNAs
CC   isolated from human, murine and rat skin cell libraries. The sequences
CC   can be used in the development of therapeutic agents useful in the
CC   treatment of skin diseases, including skin wounds, cancer, growth
CC   defects, developmental defects and inflammatory diseases. The proteins
CC   have important roles in the induction of hair growth, cell proliferation
CC   and cell-cell interaction, in maintaining tissue integrity, in wound
CC   healing and in modulating immune responses. The present sequence is a
CC   polypeptide of the invention
XX
SQ   Sequence 819 AA;

  Query Match             21.8%;  Score 832;  DB 5;  Length 819;
  Best Local Similarity   29.3%;  Pred. No. 2.4e-44;
  Matches  210;  Conservative  109;  Mismatches  264;  Indels  134;  Gaps   17;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              :| : || :: ||:||| |:           : : ||:|  |:     |||: |:||| |
Db        220 QITLEFRAEAEDGLLLYCGESE---------HGRGDFMSLALIRRSLHFRFNCGTGMAII 270

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPD-YG 119
                 | : || :|:||| |    | | : :  || | |||::  :     ||||| |  |
Db        271 ISETKIKLGAWHSVTLYRDGLNGLLQLNNGTPVTGQSQGQYSKITFRTPLYLGGAPSAYW 330

Qy        120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLN-------LTAHGISHCPT--CRDRPCQNG 170
               :   | : || |||: | : |::|   |:        |   :  | :  | :  | ||
Db        331 LVRATGTNRGFQGCVQSLAVNGKKI---DMRPWPLGKALNGADVGECSSGICDEASCING 387

Qy        171 GQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
              | |   ::  ||:|:|| || |
Db        388 GTCAAIKADSYICLCPLGFRGR-------------------------------------- 409

Qy        231 SGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFSGGK 286
                  ||:  |:| |       || | |      | |     :: |:| : |||||:|
Db        410 ---HCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSFTEFEITFRPDSGDGVLLYSYDT 466

Qy        287 SGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGG 346
              |   :||:|: |  ||:|||:: |||  |||| : | ||:|| :   |  |:| |:|:
Db        467 SS--KDFLSIIMAAGHVEFRFDCGSGTGVLRSEDTLTLGQWHDLRVSRTAKNGILQVDKQ 524

Qy        347 RPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYS 406
              : |   : |    :  :| :::|||     :   : :   | | : :: :| : : : :
Db        525 KVVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNSGILHPFSGSIQKIILNDRTIHVRHD 584

Qy        407 FLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREP 466
              |  : |:     : ||   || || :| |  |  ::| |  ||:|  |:
Db        585 F--TSGVNVENAAHPCVGAPCAHGGSCRPRKE-GYECDCPLGFEGLNCQKAI-------- 633

Qy        467 CLHGGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFL 526
                       |  : :| | |          | :    : ||  |              |:
Db        634 ---------TEAIEIPQFIGRSYLTYDNPNILK---RVSGSRSN-------------AFM 668

Qy        527 AFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFR 586
               |                    :|:   |||||:|          |||||||:|| |||
Db        669 RF--------------------KTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFS 706

Qy        587 YQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSV 646
              | |||| | ::     :|| |||| |:|:|: | | ||     :|:|||    :|  |::
Db        707 YNLGSGVASIMVNGSFSDGRWHRVKAVRDGQSGKITVDDYGARTGKSPGMMRQLNINGAL 766
```

```
Qy        647 YIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPC 703
              |:||  ::|  |  ::  |: ||: :  | :          : |   |  | |   |
Db        767 YVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DYHISLVEDAVDGKNINTC 816

RESULT 11
ABB63614
ID   ABB63614 standard; protein; 4072 AA.
XX
AC   ABB63614;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 17634.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07717.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 17634; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4072 AA;

  Query Match             21.2%;  Score 812.5;  DB 4;  Length 4072;
  Best Local Similarity   30.2%;  Pred. No. 2.6e-42;
  Matches  219;  Conservative  121;  Mismatches  304;  Indels   81;  Gaps   22;

Qy          3 KITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATIRH 62
              ::||||:: ||:||:||| |  |          |:|:  |     ||||| |     :|
Db       3394 ELTFRPENGDGLLLFNGQTRGSG----------DYIALSLKDRYAEFRFDFGGKPMLVRA 3443

Qy         63 PTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDYGAIP 122
                ||||  :||| : |    | : | :  ||   :  :   ||| |:||:|| |::  :|
Db       3444 EEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVPNWELLP 3503

Qy        123 KAGLSS--GFIGCVRELRIQGEEIVFHDLNLTAHGISHCPTCRDRPCQNGGQCHDSES-S 179
                 :|   ||:||:  | :||  :          ||: |  |   |||| | | :|::
Db       3504 ADAVSQQVGFVGCISRLTLQGRTVELIREAKYKEGITDCRPCAQGPCQNKGVCLESQTEQ 3563

Qy        180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGV 239
              :| |:|  |:||  |   :   | |  ||    | |  :     | | | ||| ||:
Db       3564 AYTCICQPGWTGRDCA-IEGTQCTPGVCGA-GRCENTEN--DMECLCPLNRSGDRCQYNE 3619

Qy        240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEFKPLA-PDGVLLFSGGKSGPVEDFVS 295
               :   ||:  |:  |    | :|    :: : :  :| :  | |:|::   : |  |:::
Db       3620 ILNEHSLNFKGNSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 3675

Qy        296 LAMVGGHLEFRYELGSGL--AVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSS 353
              | : ||| |      : |   |:|||||| | || |:   |   :| |||  |      :
Db       3676 LVLRGGHAELLINTAARLDPVVVRSAEPLPLNRWTRIEIRRRLGEGILRVGDGPERKAKA 3735

Qy        354 PGKSQGLNLHTLLYLGGVEPS-VPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFLGSQG 412
              ||  : |:| | ||:|| : | | ::   |::  | ||:  :    | ::|      :
Db       3736 PGSDRILSLKTHLYVGGYDRSTVKVNRDVNITKGFDGCISRLYNFQKPVNLLADIKDAAN 3795

Qy        413 IGQCYDSSPCERQPCQHGATCMPAGEYE------FQCLCRDGFKGDLCEHEENPCQLREP 466
              |  |             | | |  |: :            |  : :|  :   ||   :|
Db       3796 IQSC-------------GETNMIGGDEDSDNEPPVPPPTPDVHENELQPYAMAPC-ASDP 3841

Qy        467 CLHGGTCQGTR----CLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHD 522
              | :||:|        | |  ||||  ||:           ||:          : | |
Db       3842 CENGGSCSEQEDVAVCSCPFGFSGKHCQE-----------HLQLG--------FNASFRG 3882

Qy        523 DGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGH 582
              ||::      |  :| :   :: :   |:  :||| | | | ||   |:|||:  : ||:
Db       3883 DGYVELNRSHFQPALEQSYTSMGIVFTTNKPNGLLFWWGQEAGEEYTGQDFIAAAVVDGY 3942

Qy        583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALREGRRGSIQVDGEELVSGRS-PGPNVAV 640
              : :  :|   || : : |  :::|| | | | |:     ::||   | || : |    ::
Db       3943 VEYSMRLDGEEAVIRNSDIRVDNGERHIVIAKRDENTAILEVD-RMLHSGETRPTSKKSM 4001

Qy        641 NAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANT 700
                 |:|::|||||:   || |:   : ||:  :|:     |      ::|   |  | |
Db       4002 KLPGNVFVGGAPDLEVFTGFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 4054

Qy        701 RPCPS 705
                ||:
Db       4055 NVCPA 4059

RESULT 12
ABB61948
ID   ABB61948 standard; protein; 1298 AA.
XX
AC   ABB61948;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 12636.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06051.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
```

```
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 12636; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1298 AA;

  Query Match             17.4%;  Score 664.5;  DB 4;  Length 1298;
  Best Local Similarity   22.0%;  Pred. No. 1.7e-33;
  Matches  221;  Conservative  111;  Mismatches  295;  Indels  379;  Gaps   23;

Qy          1 EIKITFRPDSADGMLLYNGQKRVPGSPTNLANRQPDFISFGLVGGRPEFRFDAGSGMATI 60
              :::| |||:| ||::| :|::          :   ||::  |  |  || || |||: ::
Db        361 QLRIEFRPESFDGIILLSGER---------DDLTGDFMALLLNKGFVEFWFDCGSGVGSV 411

Qy         61 RHPTPLALGHFHTVTLLRSLTQGSLIVGDLAPVNGTSQGKFQGLDLNEELYLGGYPDY-G 119
              |    : |  :::| : |      |::     | | | | |  :   | ::|||  :  |
Db        412 RSRETILLNEWNSVIIYRHRWDAWLVLNHGTKVQGRSNGLFSRITFREPVFLGGIGNITG 471

Qy        120 AIPKAGLSSGFIGCVRELRIQGEEIVFHDLNLTAH---------GISHCPT--CRDRPCQ 168
                 :  |: || ||:|  |    |   ||   | |          |  | |  |   |||
Db        472 LAKRLPLAEGFAGCIR--RFVANE---HDYKFTEHPLGDVINGFDIQDCSTDKCVRYPCQ 526

Qy        169 NGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
              :||:|  |:  : :|:|| || |  ||
Db        527 HGGKCLPSDQGA-ICLCPIGFVGDLCE--------------------------------- 552

Qy        229 GRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELRLDVEFKPLAPDGVLLFSGGKSG 288
                  :| :    :  |: :|:      |   :    | | |  ||   ||::|:|| :
Db        553 ----IRMD----LQVPAFNGSSFLRYAPLGDSALIWLELKVTLKPEQADGLILYSGPEHR 604

Qy        289 PVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRP 348
                 ||::| :  | :|| ::|||| |::||   |:||:|| :   |  :   |:|:  :
Db        605 --GDFIALYLNDGFVEFAFDLGSGPALVRSEHSLSLGQWHTIKISRTARLAVLKVDKHQE 662

Qy        349 VLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRLDLTYSFL 408
              ||  |      |:|   |::|||     |         | ||:  : :||  | :    |
Db        663 VLTISSNGFWHLSLDQNLFVGGVNHVDRLPLDLKYKPFFVGCIQRIDINGHSLGIVAEAL 722

Qy        409 GSQGIGQCYDSSPCERQPCQHGATCMPAGE-YEFQCLCRDGFKGDLCEH--EENPCQLRE 465
              |   || |     |  :||   | |:|  | ||    ||     : |    |  | || |
Db        723 GGSNIGNC--PHACVARPCGPLAECVPQMESYE----CRCSIHNERCNKAAEVPPEQLPE 776

Qy        466 PCLHGGTCQGTR------------------------------------------------ 477
                ||      |:
Db        777 LALHKSKVLETKDNGEAAKKVSGLAHKKHSKKHRNLHKPTPATSTTTSTTSTTTTTTEAP 836

Qy        478 --------------------------------------------------------CLCL 481

Db        837 SERTEEATAGALSNEEIEDDIIFRLVQQQQQQKELKKQHQQTTTTAPATSTTSSGPSKAK 896

Qy        482 PGFSG------------------------------------------------------- 486
              |  ||
Db        897 PRLSGKHHASKHEHHLKPNAAFTRKLSRLPTHYESFQTNPDSDILTFEDNNDWVTSLQQQ 956

Qy        487 ----------------------PRCQQG-------------------------------- 492
                                    ||
Db        957 EYGDAMAASQVPLAFEDASPGTPRSSDNNEDDENAFVFDESLFDASDGTEEYQRKQLAQD 1016

Qy        493 -----------SGHGIAESDWHLEGS----GGNDAPGQYGAYFHDDGFL----------- 526
                         | |  ||  :  :||      |:   ||   ::||  |
Db       1017 MKRIMSNSNAHSSHKKAEVQFPPQGSQEVGTANEDTSQYSDDYNDDELLTPVMQGGEEVK 1076

Qy        527 ----------------------------------------AFPG--------------HV 532
                                                       | |               |
Db       1077 LEQHTSSTPQTHTDWSLLKKFDLSAEHQSQVQGVRKNFGACFAGSDSYFHYNDADTMSQV 1136

Qy        533 FSRSLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSG 592
               | |:      : | ::| : :|::|| | : |   :  |::|||:: |:| ||| ||||
Db       1137 ISYSI-----DLNLRIKTHSENGVILWTGRQ-GTTEEHDDYLSLGIEQGYLHFRYDLGSG 1190

Qy        593 EARL-VSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGA 651
              |  :  :   ::|| |||| |:|  : | ::||| : |: |:||    :|    :|:||
Db       1191 EVDIRFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKTVTLRAPGKLRQLNTDTGLYVGGM 1250

Qy        652 PDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAG 697
              |||   |  |: ||| ||:  :||         |  |  :|  : |
Db       1251 PDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDPNTLGTEHNVETG 1296

RESULT 13
AAW26609
ID   AAW26609 standard; protein; 492 AA.
XX
AC   AAW26609;
XX
DT   27-JAN-1998  (first entry)
XX
DE   Human agrin.
XX
KW   Agrin; human; receptor tyrosine kinase; muscle specific kinase; MuSK;
KW   ligand; muscle atrophy; muscular dystrophy; myopathy; diagnosis;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= Sig_peptide
FT   Protein         24. .492
FT                   /note= "Claim 2"
FT   Protein         60. .492
FT                   /note= "Claim 2"
FT   Protein         76. .492
FT                   /note= "Claim 2"
FT   Protein         126. .492
FT                   /note= "Claim 2"
FT   Protein         178. .492
FT                   /note= "Claim 2"
FT   Peptide         187. .190
FT                   /label= Y-insert
FT                   /note= "sequence absent from rat agrin"
FT   Protein         222. .492
FT                   /note= "Claim 2"
FT   Protein         260. .492
FT                   /note= "Claim 2"
FT   Protein         300. .492
FT                   /note= "Claim 2"
FT   Peptide         324. .331
FT                   /label= Z-insert
FT                   /note= "sequence absent from rat agrin"
XX
PN   WO9721811-A2.
XX
PD   19-JUN-1997.
XX
PF   13-DEC-1996;    96WO-US020696.
XX
PR   15-DEC-1995;    95US-0008657P.
PR   10-MAY-1996;    96US-00644271.
XX
PA   (REGE-) REGENERON PHARM INC.
XX
PI   Valenzuela DM,  Glass DJ,  Bowen DC,  Yancopoulos GD;
XX
DR   WPI; 1997-332783/30.
```

```
DR   N-PSDB; AAT90471.
XX
PT   Nucleotide sequences encoding human agrin and muscle specific kinase and
PT   related receptor - used in diagnosis and treatment of disorder with
PT   muscle atrophy.
XX
PS   Claim 2; Fig 15; 120pp; English.
XX
CC   This polypeptide comprises the amino acid sequence of human agrin deduced
CC   from an isolated foetal brain cDNA clone (see AAT90471). Full-length
CC   agrin, or active portions of it, can be expressed using a claimed host-
CC   vector system. The agrin polypeptides can be used in claimed methods of
CC   treating diseases or disorders that affect muscle, especially atrophy
CC   resulting from denervation due to nerve trauma, degenerative, metabolic
CC   or inflammatory neuropathy, peripheral neuropathy or damage to nerves
CC   caused by environmental toxins or drugs. In particular the muscle atrophy
CC   is due to motor neuronopathy, chronic disuse, metabolic stress or
CC   nutritional insufficiency, muscular dystrophy syndrome, congenital
CC   myopathy or acquired (toxic or inflammatory) myopathy. The polypeptides
CC   can also be used in a claimed method of promoting the growth,
CC   differentiation or survival of cells that express muscle-specific (MuSK)
CC   receptors (see AAW26610-11), and to raise antibodies
XX
SQ   Sequence 492 AA;

  Query Match             16.7%;  Score 640.5;  DB 2;  Length 492;
  Best Local Similarity   33.0%;  Pred. No. 1.9e-32;
  Matches  183;  Conservative   59;  Mismatches  173;  Indels  139;  Gaps   18;

Qy        202 CHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNT 261
              | |  |   | |   |:| |  | | ||| |  |:     |     |:| :||      :
Db         27 CQPNPCHGAAPCRVLPEG-GAQCECPLGREGTFCQ-----TASGQDGSGPFLA-DFNGFS 79

Qy        262 HHELR--------------LDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRY 307
              | |||              |:| |    | |:||::| |:    ||||||:    |||||
Db         80 HLELRGLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFRY 139

Qy        308 ELGSGLAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQG-----LNL 362
              :|| | ||:|| ||: || | ||| ||  : |:|||  |  ||  || ||:      |||
Db        140 DLGKGAAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESP-KSRKVPHTVLNL 198

Qy        363 HTLLYLGGVEPSVPLSPATNMSAHFRGCVGEVSVNGKRL----------DLTYSFLGSQG 412
                 ||:||      |: |  :|: | | :   ||: |::|          |:| || |
Db        199 KEPLYVGGAPDFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG--- 254

Qy        413 IGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLH 469
                      || |    || :||:|:|                            ||
Db        255 -------HPCTRASGHPCLNGASCVP----------------------------REAAY- 278

Qy        470 GGTCQGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFP 529
                      |||  ||||| |::|           :| | |            |   |||
Db        279 -------VCLCPGGFSGPHCEKGL----------VEKSAG------------DVDTLAFD 309

Qy        530 GHVFSRSLPEVPET----------------IELEVRTSTASGLLLWQGVEVGEAGQGKDF 573
              |  |   |  | |:                 || :||     |:||     |:| :  |:
Db        310 GRTFVEYLNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLWS----GKATERADY 365

Qy        574 ISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRS 633
              ::| : ||||   | |||    | |  |:|   | || | || | ||:||  |  |:| |
Db        366 VALAIVDGHLQLSYNLGSQPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSS 425

Qy        634 PGPNVAVNAKGSVYIGGAPD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQ 691
              |     ::  |::::|| |:  |       : :|  ||::::|:           || |
Db        426 PLGATQLDTDGALWLGGLPELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLL 477

Qy        692 HRAQAGANTRPCPS 705
                |      ||||:
Db        478 EDAVTKPELRPCPT 491

RESULT 14
ADC39166
ID   ADC39166 standard; protein; 1566 AA.
```

```
XX
AC   ADC39166;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Novel human NOVX polypeptide  SEQ ID NO: 108.
XX
KW   antidiabetic; cytostatic; immunomodulator; anorectic; antilipemic;
KW   nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
KW   antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
KW   hemostatic; osteopathic; gene therapy.; NOVX; diabetes; obesity; cancer;
KW   lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
KW   wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
KW   cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
KW   hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
KW   Albright hereditary osteodystrophy.
XX
OS   Homo sapiens.
XX
PN   WO2003010327-A2.
XX
PD   06-FEB-2003.
XX
PF   02-MAY-2002; 2002WO-US014199.
XX
PR   02-MAY-2001; 2001US-0288063P.
PR   03-MAY-2001; 2001US-0288395P.
PR   07-MAY-2001; 2001US-0289087P.
PR   09-MAY-2001; 2001US-0289817P.
PR   09-MAY-2001; 2001US-0289818P.
PR   11-MAY-2001; 2001US-0290194P.
PR   14-MAY-2001; 2001US-0290753P.
PR   15-MAY-2001; 2001US-0291181P.
PR   16-MAY-2001; 2001US-0291243P.
PR   18-MAY-2001; 2001US-0292001P.
PR   21-MAY-2001; 2001US-0292374P.
PR   22-MAY-2001; 2001US-0292587P.
PR   23-MAY-2001; 2001US-0293107P.
PR   25-MAY-2001; 2001US-0293747P.
PR   29-MAY-2001; 2001US-0294109P.
PR   29-MAY-2001; 2001US-0294110P.
PR   30-MAY-2001; 2001US-0294434P.
PR   31-MAY-2001; 2001US-0294827P.
PR   12-JUL-2001; 2001US-0304879P.
PR   31-JUL-2001; 2001US-0308901P.
PR   14-AUG-2001; 2001US-0312270P.
PR   17-AUG-2001; 2001US-0313416P.
PR   10-SEP-2001; 2001US-0318463P.
PR   27-SEP-2001; 2001US-0325683P.
PR   18-OCT-2001; 2001US-0330292P.
PR   28-NOV-2001; 2001US-0333873P.
PR   03-DEC-2001; 2001US-0336909P.
PR   03-DEC-2001; 2001US-0337552P.
PR   21-FEB-2002; 2002US-0359245P.
PR   01-MAY-2002; 2002US-00136826.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Miller CE,  Kekuda R,  Malyankar UM,  Li L,  Pena CEA,  Spytek KA;
PI   Gorman L,  Guo X,  Fernandes ER,  Smithson G,  Stone DJ,  Zerhusen BD;
PI   Patturajan M,  Anderson DW,  Mezes PS,  Peyman JA,  Macdougall JR;
PI   Padigaru M,  Rastelli L,  Shenoy SG,  Gerlach VL,  Shimkets RA,  Zhong M;
PI   Edinger SR,  Ellerman K;
XX
DR   WPI; 2003-239445/23.
DR   N-PSDB; ADC39165.
XX
PT   New NOVX polypeptides and polynucleotides, useful in gene therapy,
PT   particularly for treating or preventing a syndrome associated with a
PT   human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
PT   hypertension or hemophilia.
XX
PS   Claim 1; SEQ ID NO 108; 748pp; English.
```

```
XX
CC   The invention relates to new isolated NOVX polypeptides, the genes
CC   encoding them or sequences having at least 95% identity to the amino acid
CC   or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
CC   particularly in the manufacture of a medicament for treating a syndrome
CC   associated with a human disease, which includes a pathology associated
CC   with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC   treating, preventing or alleviating pathology associated with NOVX
CC   polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC   polypeptide are especially useful for treating or preventing e.g.
CC   diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC   cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
CC   Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC   disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
CC   graft-versus-host disease or Albright hereditary osteodystrophy. The DNA
CC   encoding the protein is useful in gene therapy for treating the above
CC   conditions. These are also useful in developing powerful assay system for
CC   functional analysis of various human disorders, as well as in diagnostic
CC   applications. This sequence represents one of the NOVX proteins of the
CC   invention.
XX
SQ   Sequence 1566 AA;

  Query Match             16.6%;  Score 635.5;  DB 7;  Length 1566;
  Best Local Similarity   33.4%;  Pred. No. 1.5e-31;
  Matches  179;  Conservative   58;  Mismatches  174;  Indels  125;  Gaps   16;

Qy        207 CGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSLSGAGSYLALPALTNTHHELR 266
              |   | |   |:| |  | | ||| |  |:     |     |:| :||      :| |||
Db       1118 CHGAAPCRVLPEG-GAQCECPLGREGTFCQ-----TASGQDGSGPFLA-DFNGFSHLELR 1170

Qy        267 --------------LDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSG 312
                            |:| |    | |:||::| |:    ||||||:    |||||:|| |
Db       1171 GLHTFARDLGEKMALEVVFLARGPSGLLLYNGQKTDGKGDFVSLALRDRRLEFRYDLGKG 1230

Qy        313 LAVLRSAEPLALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVE 372
               ||:|| ||: || | ||| ||  : |:|||  |  ||  ||     |||   ||:||
Db       1231 AAVIRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAP 1290

Qy        373 PSVPLSPATNMSAHFRGCVGEVSVNGKRL----------DLTYSFLGSQGIGQCYDSSPC 422
                  |: |  :|: | | :  ||: |::|          |:| || |           ||
Db       1291 DFSKLARAAAVSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG----------HPC 1339

Qy        423 ER---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCL 479
               |    || :||:|:|                            ||            ||
Db       1340 TRASGHPCLNGASCVP----------------------------REAAY--------VCL 1363

Qy        480 CLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPE 539
              |  ||||| |::|           :| | |            |   ||| |  |   |
Db       1364 CPGGFSGPHCEKGL-----------VEKSAG-------------DVDTLAFDGRTFVEYLNA 1401

Qy        540 VPET---------IELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGS 591
              | |:         || :||    ||:||     |:| :  |:::| : ||||   | |||
Db       1402 VTESEKALQSNHFELSLRTEATQGLVLWS----GKATERADYVALAIVDGHLQLSYNLGS 1457

Qy        592 GEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGA 651
                  | |  |:|   | || | || | ||:||  |  |:| ||     ::  |::::||
Db       1458 QPVVLRSTVPVNTNRWLRVVAHREQREGSLQVGNEAPVTGSSPLGATQLDTDGALWLGGL 1517

Qy        652 PD--VATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 705
              |:  |       : :|  ||::::|:           || |   |      ||||:
Db       1518 PELPVGPALPKAYGTGFVGCLRDVVVGR--------HPLHLLEDAVTKPELRPCPT 1565

RESULT 15
ADB64904
ID   ADB64904 standard; protein; 463 AA.
XX
AC   ADB64904;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human protein encoded by clone PLACE60113340.
XX
KW   Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW   cell regeneration; membrane protein; signal transduction-related protein;
KW   transcription-related protein; osteoporosis; neurological disease;
KW   cancer; tumour.
XX
OS   Homo sapiens.
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   N-PSDB; ADB62934.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a protein of the invention. Note: Some of the
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 463 AA;

  Query Match             16.4%;  Score 628;  DB 7;  Length 463;
  Best Local Similarity   28.7%;  Pred. No. 1.1e-31;
  Matches  162;  Conservative   80;  Mismatches  198;  Indels  124;  Gaps   16;

Qy        151 LTAHGISHCPT--CRDRPCQNGGQCHDSESSSYVCVCPAGFTGSRCEHSQALHCHPEACG 208
              |:   :  | :  | :  | :|| |   :: ||:|:|| ||
Db         10 LSGADVGECSSGICDEASCIHGGTCTAIKADSYICLCPLGF------------------- 50

Qy        209 PDATCVNRPDGRGYTCRCHLGRSGLRCEEGVTVTTPSL-SGAGSYLALPALTNTHHEL-- 265
                                     |  ||:  |:| |       || | |      | |
Db         51 ----------------------KGRHCEDAFTLTIPQFRESLRSYAATPWPLEPQHYLSF 88

Qy        266 -RLDVEFKPLAPDGVLLFS---GGKSGPVEDFVSLAMVGGHLEFRYELGSGLAVLRSAEP 321
```

```
                 :: |:| : |||||:|   | |     ||:|: : |||:|||:: |||  |||| :|
Db         89 MEFEITFRPDSGDGVLLYSYDTGSK-----DFLSINLAGGHVEFRFDCGSGTGVLRSEDP 143

Qy        322 LALGRWHRVSAERLNKDGSLRVNGGRPVLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPAT 381
              | || || :   |  |:| |:|:  : |   : |    :  :| :::|||     :   :
Db        144 LTLGNWHELRVSRTAKNGILQVDKQKIVEGMAEGGFTQIKCNTDIFIGGVPNYDDVKKNS 203

Qy        382 NMSAHFRGCVGEVSVNGKRLDLTYSFLGSQGIGQCYDSSPCERQPCQHGATCMPAGEYEF 441
               :   | | : :: :| : : : : |  : |:     : || | || || :| |  |  :
Db        204 GVLKPFSGSIQKIILNDRTIHVKHDF--TSGVNVENAAHPCVRAPCAHGGSCRPRKE-GY 260

Qy        442 QCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGT--RCLCLPGFSGPRCQQGSGHGIAE 499
               | |  ||:|                ||   ||      : :| | |
Db        261 DCDCPLGFEG----------------LH---CQKAIIEAIEIPQFIG------------- 288

Qy        500 SDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLW 559
                              |  | : |      |   |||       : :  :|:   |||||
Db        289 --------------RSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLW 325

Qy        560 QGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRG 619
              :|          |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: |
Db        326 RG--DSPMRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSG 383

Qy        620 SIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSAR 679
               | ||     :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        384 KITVDDYGARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-- 441

Qy        680 PGAPPPQPLDLQHRAQAGANTRPC 703
                      : |   |  | |   |
Db        442 -----DYHISLVEDAVDGKNINTC 460

Search completed: March  9, 2004, 17:19:02
Job time : 91.8465 secs
```

```
                      GenCore version 5.1.6
             Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:11:52 ; Search time 6.86906 Seconds
                                            (without alignments)
                                            2152.832 Million cell updates/sec

Title:          US-10-006-011A-9
Perfect score:  1566
Sequence:       1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:       141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                    %
Result            Query
  No.    Score    Match Length DB  ID                        Description
---------------------------------------------------------------------------------
    1     1566   100.0    4391  1  PGBM_HUMAN                P98160 homo sapien
    2   1370.5    87.5    3707  1  PGBM_MOUSE                Q05793 mus musculu
    3      349    22.3    3375  1  UN52_CAEEL                Q06561 caenorhabdi
    4    333.5    21.3    1959  1  AGRI_RAT                  P25304 rattus norv
    5    314.5    20.1    1328  1  AGRI_DISOM                Q90404 discopyge o
    6    301.5    19.3    1955  1  AGRI_CHICK                P31696 gallus gall
    7      269    17.2    5147  1  FAT_DROME                 P33450 drosophila
    8    253.5    16.2    1529  1  SLT2_HUMAN                O94813 homo sapien
    9      242    15.5    4705  1  FAT2_DROME                Q9vw71 drosophila
   10      225    14.4    3579  1  STAN_DROME                Q9v5n8 drosophila
   11    223.5    14.3    3110  1  LMA2_HUMAN                P24043 homo sapien
   12      214    13.7    1578  1  NX3A_RAT                  Q07310 rattus norv
   13    211.5    13.5    2923  1  CLR2_HUMAN                Q9hcu4 homo sapien
   14    207.5    13.3    2144  1  CLR2_RAT                  Q9qyp2 rattus norv
   15    207.5    13.3    3106  1  LMA2_MOUSE                Q60675 mus musculu
   16    206.5    13.2    3312  1  CLR3_HUMAN                Q9nyq7 homo sapien
   17    204.5    13.1    1712  1  NX2A_HUMAN                Q9p2s2 homo sapien
   18    204.5    13.1    3313  1  CLR3_RAT                  O88278 rattus norv
   19      200    12.8    1541  1  NX3A_HUMAN                Q9y4c0 homo sapien
   20      199    12.7    2321  1  NTC3_HUMAN                Q9um47 homo sapien
   21      199    12.7    3014  1  CLR1_HUMAN                Q9nyq6 homo sapien
   22    197.5    12.6    1530  1  NX1A_BOVIN                Q28146 bos taurus
   23    194.5    12.4    1715  1  NX2A_RAT                  Q63374 rattus norv
   24      193    12.3    1363  1  NX1A_CHICK                Q9ddd0 gallus gall
   25      193    12.3    1477  1  NX1A_HUMAN                Q9ulb1 homo sapien
   26      193    12.3    1514  1  NX1A_RAT                  Q63372 rattus norv
   27      191    12.2    3034  1  CLR1_MOUSE                O35161 mus musculu
   28    190.5    12.2     618  1  DLL3_HUMAN                Q9nyj7 homo sapien
   29    189.5    12.1    3301  1  CLR3_MOUSE                Q91zi0 mus musculu
   30    188.5    12.0    3075  1  LMA1_HUMAN                P25391 homo sapien
   31      188    12.0    2215  1  CDN2_DROME                Q9vjb6 drosophila
   32    185.5    11.8    3672  1  LML2_CAEEL                Q21313 caenorhabdi
   33      184    11.7    3097  1  CADN_DROME                O15943 drosophila
```

```
   34      183    11.7    1504  1  SLIT_DROME                P24014 drosophila
   35    182.5    11.7    2003  1  NTC4_HUMAN                Q99466 homo sapien
   36      182    11.6    2318  1  NTC3_MOUSE                Q61982 mus musculu
   37    181.5    11.6    2531  1  NTC1_MOUSE                Q01705 mus musculu
   38    181.5    11.6    2531  1  NTC1_RAT                  Q07008 rattus norv
   39    180.5    11.5     589  1  DLL3_RAT                  O88671 rattus norv
   40    180.5    11.5     592  1  DLL3_MOUSE                O88516 mus musculu
   41    180.5    11.5    1247  1  JAG2_MOUSE                Q9qye5 mus musculu
   42    180.5    11.5    1321  1  PGCN_HUMAN                O14594 homo sapien
   43      180    11.5    1238  1  JAG2_HUMAN                Q9y219 homo sapien
   44    179.5    11.5    1202  1  JAG2_RAT                  P97607 rattus norv
   45    179.5    11.5    1242  1  JAG1_BRARE                Q90y57 brachydanio

                                  ALIGNMENTS

RESULT 1
PGBM_HUMAN
ID   PGBM_HUMAN     STANDARD;      PRT;  4391 AA.
AC   P98160; Q16287; Q9H3V5;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Skin;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX   MEDLINE=20553141; PubMed=11101850;
RA   Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA   Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA   Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA   Hentati F., Fontaine B.;
RT   "Perlecan, the major proteoglycan of basement membranes, is altered in
RT   patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL   Nat. Genet. 26:480-483(2000).
RN   [4]
RP   SEQUENCE OF 1016-1470 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [5]
RP   SEQUENCE OF 890-1396 FROM N.A.
RC   TISSUE=Fibrosarcoma;
```

```
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [6]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN   [7]
RP   CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX   MEDLINE=22660472; PubMed=12754519;
RA   Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT   "Identification and quantification of N-linked glycoproteins using
RT   hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL   Nat. Biotechnol. 21:660-666(2003).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC       syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC       characterized by permanent myotonia (prolonged failure of muscle
CC       relaxation) and skeletal dysplasia, resulting in reduced stature,
CC       kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; AL445795; CAC18534.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   PIR; A38096; A38096.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
DR   MIM; 142461; -.
DR   MIM; 255800; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
```

```
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00180; EGF_Lam; 12.
DR   SMART; SM00409; IG; 22.
DR   SMART; SM00408; IGc2; 21.
DR   SMART; SM00406; IGv; 7.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 22.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; Disease mutation.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4391       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      198    235       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      284    320       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      324    360       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      367    404       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      405    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1955       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1956   2051       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2052   2151       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2152   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2340       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2341   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2533       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2534   2629       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2630   2726       IG-LIKE C2-TYPE 12.
```

```
  Query Match             100.0%;  Score 1566;  DB 1;  Length 4391;
  Best Local Similarity   100.0%;  Pred. No. 2e-103;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4108 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 4167

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4168 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 4227

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4228 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 4287

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4288 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 4347

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||||||||||||||||||||||||||||||||||||||
Db       4348 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 4391


RESULT 2
PGBM_MOUSE
ID   PGBM_MOUSE     STANDARD;      PRT;  3707 AA.
AC   Q05793;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=92078153; PubMed=1744087;
RA   Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "The complete sequence of perlecan, a basement membrane heparan
RT   sulfate proteoglycan, reveals extensive similarity with laminin A
RT   chain, low density lipoprotein-receptor, and the neural cell adhesion
RT   molecule.";
RL   J. Biol. Chem. 266:22939-22947(1991).
RN   [2]
RP   SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89034110; PubMed=2972708;
RA   Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA   Yamada Y., Hassell J.R.;
RT   "Identification of cDNA clones encoding different domains of the
RT   basement membrane heparan sulfate proteoglycan.";
RL   J. Biol. Chem. 263:16379-16387(1988).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
```

```
CC   -!- SIMILARITY: Contains 3 laminin IV domains.
CC   -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M77174; AAA39911.1; -.
DR   EMBL; J04054; AAA39899.1; -.
DR   EMBL; J04055; AAA39912.1; -.
DR   PIR; S18252; S18252.
DR   PDB; 1GL4; 28-NOV-01.
DR   MGD; MGI:96257; Hspg2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   GO; GO:0008104; P:protein localization; IMP.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 15.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 7.
DR   SMART; SM00408; IGc2; 14.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50835; IG_LIKE; 15.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain; 3D-structure.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   3707       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
FT   DOMAIN      195    234       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      281    319       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      320    359       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      360    403       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      404    504       IG-LIKE C2-TYPE 1.
FT   DOMAIN      521    530       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN      531    730       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN      731    763       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN      764    813       LAMININ EGF-LIKE 2.
FT   DOMAIN      814    871       LAMININ EGF-LIKE 3.
```

```
FT   DOMAIN      879    923       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN      924    933       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      934   1125       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN     1126   1158       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN     1159   1208       LAMININ EGF-LIKE 6.
FT   DOMAIN     1209   1265       LAMININ EGF-LIKE 7.
FT   DOMAIN     1275   1324       LAMININ EGF-LIKE 8.
FT   DOMAIN     1325   1334       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN     1335   1529       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN     1530   1562       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN     1563   1612       LAMININ EGF-LIKE 10.
FT   DOMAIN     1613   1670       LAMININ EGF-LIKE 11.
FT   DOMAIN     1677   1771       IG-LIKE C2-TYPE 2.
FT   DOMAIN     1772   1865       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1866   1954       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1955   2049       IG-LIKE C2-TYPE 5.
FT   DOMAIN     2050   2148       IG-LIKE C2-TYPE 6.
FT   DOMAIN     2149   2244       IG-LIKE C2-TYPE 7.
FT   DOMAIN     2245   2343       IG-LIKE C2-TYPE 8.
FT   DOMAIN     2344   2436       IG-LIKE C2-TYPE 9.
FT   DOMAIN     2437   2532       IG-LIKE C2-TYPE 10.
FT   DOMAIN     2533   2619       IG-LIKE C2-TYPE 11.
FT   DOMAIN     2620   2720       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2721   2809       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2810   2895       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2896   2980       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2984   3162       LAMININ G-LIKE 1.
FT   DOMAIN     3163   3241       EGF-LIKE.
FT   DOMAIN     3245   3425       LAMININ G-LIKE 2.
FT   DOMAIN     3518   3705       LAMININ G-LIKE 3.
FT   SITE         65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE         71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE         76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE       3615   3617       MEDIATES MOTOR NEURON ATTACHMENT
FT                                (POTENTIAL).
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    206    225       BY SIMILARITY.
FT   DISULFID    219    234       BY SIMILARITY.
FT   DISULFID    285    297       BY SIMILARITY.
FT   DISULFID    292    310       BY SIMILARITY.
FT   DISULFID    304    319       BY SIMILARITY.
FT   DISULFID    325    337       BY SIMILARITY.
FT   DISULFID    332    350       BY SIMILARITY.
FT   DISULFID    344    359       BY SIMILARITY.
FT   DISULFID    368    381       BY SIMILARITY.
FT   DISULFID    375    394       BY SIMILARITY.
FT   DISULFID    388    403       BY SIMILARITY.
FT   DISULFID    428    479       BY SIMILARITY.
FT   DISULFID    764    773       BY SIMILARITY.
FT   DISULFID    766    780       BY SIMILARITY.
FT   DISULFID    783    792       BY SIMILARITY.
FT   DISULFID    795    811       BY SIMILARITY.
FT   DISULFID    814    829       BY SIMILARITY.
FT   DISULFID    816    839       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    869       BY SIMILARITY.
FT   DISULFID   1159   1168       BY SIMILARITY.
FT   DISULFID   1161   1175       BY SIMILARITY.
FT   DISULFID   1178   1187       BY SIMILARITY.
FT   DISULFID   1190   1206       BY SIMILARITY.
FT   DISULFID   1209   1224       BY SIMILARITY.
FT   DISULFID   1211   1234       BY SIMILARITY.
FT   DISULFID   1237   1246       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1275   1287       BY SIMILARITY.
FT   DISULFID   1277   1293       BY SIMILARITY.
FT   DISULFID   1295   1304       BY SIMILARITY.
FT   DISULFID   1307   1322       BY SIMILARITY.
FT   DISULFID   1563   1572       BY SIMILARITY.
FT   DISULFID   1565   1579       BY SIMILARITY.
FT   DISULFID   1582   1591       BY SIMILARITY.
FT   DISULFID   1594   1610       BY SIMILARITY.
FT   DISULFID   1613   1628       BY SIMILARITY.
```

```
FT   DISULFID   1615   1638       BY SIMILARITY.
FT   DISULFID   1641   1650       BY SIMILARITY.
FT   DISULFID   1653   1668       BY SIMILARITY.
FT   DISULFID   1792   1839       BY SIMILARITY.
FT   DISULFID   1886   1932       BY SIMILARITY.
FT   DISULFID   1976   2021       BY SIMILARITY.
FT   DISULFID   2073   2118       BY SIMILARITY.
FT   DISULFID   2170   2215       BY SIMILARITY.
FT   DISULFID   2268   2313       BY SIMILARITY.

  Query Match             87.5%;  Score 1370.5;  DB 1;  Length 3707;
  Best Local Similarity   86.3%;  Pred. No. 1.4e-89;
  Matches  245;  Conservative   20;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQGTRCLCL 60
              ||||||::||||||||||||||||:||||||||||||||||| ||||:||||:| |||||
Db       3425 CERQPCRNGATCMPAGEYEFQCLCQDGFKGDLCEHEENPCQLHEPCLNGGTCRGARCLCL 3484

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVP 120
              |||||||||||:|:|: ||||| ||||||||||||||||:|:|||  ||: |||||||||
Db       3485 PGFSGPRCQQGAGYGVVESDWHPEGSGGNDAPGQYGAYFYDNGFLGLPGNSFSRSLPEVP 3544

Qy        121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP 180
              |||| ||||||| |||||||| | || : ||||||||||||||| |||||||||||| ||
Db       3545 ETIEFEVRTSTADGLLLWQGV-VREASRSKDFISLGLQDGHLVFSYQLGSGEARLVSGDP 3603

Qy        181 INDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGR 240
              ||||||||:||||||:|||||||||:||:||||||||||| |   :|||||||||||| |:
Db       3604 INDGEWHRITALREGQRGSIQVDGEDLVTGRSPGPNVAVNTKDIIYIGGAPDVATLTRGK 3663

Qy        241 FSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
              ||||||||:||||||:||||||||||||||||||||||||||||
Db       3664 FSSGITGCIKNLVLHTARPGAPPPQPLDLQHRAQAGANTRPCPS 3707


RESULT 3
UN52_CAEEL
ID   UN52_CAEEL     STANDARD;      PRT;  3375 AA.
AC   Q06561; O18261; O18263; Q9XTD2; Q9XTI5;
DT   01-JUN-1994 (Rel. 29, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Basement membrane proteoglycan precursor (Perlecan homolog)
DE   (Uncoordinated protein 52).
GN   UNC-52 OR ZC101.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.
RX   MEDLINE=93339574; PubMed=8393416;
RA   Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
RT   "Products of the unc-52 gene in Caenorhabditis elegans are homologous
RT   to the core protein of the mammalian basement membrane heparan
RT   sulfate proteoglycan.";
RL   Genes Dev. 7:1471-1484(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Percy C.M., Baynes C.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RA   Durbin R.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probable role in myofilament assembly and/or attachment
CC       of the myofilament lattice to the cell membrane. May be an
CC       extracellular anchor for integrin receptors in muscle.
CC   -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=e;
```

```
CC        IsoId=Q06561-1; Sequence=Displayed;
CC          Note=No experimental confirmation available;
CC      Name=a;
CC        IsoId=Q06561-2; Sequence=VSP_007195, VSP_007196;
CC      Name=b;
CC        IsoId=Q06561-3; Sequence=VSP_007191, VSP_007192;
CC          Note=No experimental confirmation available;
CC      Name=c;
CC        IsoId=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
CC                        VSP_007196;
CC          Note=No experimental confirmation available;
CC  -!- TISSUE SPECIFICITY: Found in the basement membrane of all
CC      contractile tissues. It is concentrated over muscle dense bodies
CC      and M-lines which are associated with beta-integrin.
CC  -!- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
CC  -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC  -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC  -!- SIMILARITY: Contains 7 laminin EGF-like domains.
CC  -!- SIMILARITY: Contains 3 laminin G-like domains.
CC  -!- SIMILARITY: Contains 2 laminin IV domains.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC  the European Bioinformatics Institute.  There are no  restrictions on  its
CC  use  by  non-profit  institutions as long  as its content  is  in  no  way
CC  modified and this statement is not removed.  Usage  by  and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR  EMBL; L13458; AAA28156.1; -.
DR  EMBL; Z93375; CAB07567.1; -.
DR  EMBL; Z93395; CAB07567.1; JOINED.
DR  EMBL; Z93375; CAB07568.1; -.
DR  EMBL; Z93395; CAB07568.1; JOINED.
DR  EMBL; Z93375; CAB07569.1; -.
DR  EMBL; Z93395; CAB07569.1; JOINED.
DR  EMBL; Z93395; CAB07704.1; -.
DR  EMBL; Z93395; CAB07706.1; -.
DR  EMBL; Z93375; CAB07706.1; JOINED.
DR  EMBL; Z93395; CAB07707.1; -.
DR  EMBL; Z93375; CAB07707.1; JOINED.
DR  EMBL; Z93395; CAB07708.1; -.
DR  EMBL; Z93375; CAB07708.1; JOINED.
DR  HSSP; P01130; 1LDR.
DR  WormPep; ZC101.2a; CE15028.
DR  WormPep; ZC101.2b; CE15030.
DR  WormPep; ZC101.2c; CE15034.
DR  WormPep; ZC101.2e; CE18424.
DR  GO; GO:0005578; C:extracellular matrix; IEP.
DR  GO; GO:0030239; P:myofibril assembly; IEP.
DR  InterPro; IPR008985; ConA_like_lec_gl.
DR  InterPro; IPR001881; EGF_Ca.
DR  InterPro; IPR006209; EGF_like.
DR  InterPro; IPR006210; IEGF.
DR  InterPro; IPR007110; Ig-like.
DR  InterPro; IPR003599; Ig.
DR  InterPro; IPR003598; Ig_c2.
DR  InterPro; IPR000034; Laminin_B.
DR  InterPro; IPR002049; Laminin_EGF.
DR  InterPro; IPR001791; Laminin_G.
DR  InterPro; IPR002172; LDL_receptor_A.
DR  Pfam; PF00047; ig; 16.
DR  Pfam; PF00052; laminin_B; 2.
DR  Pfam; PF00053; laminin_EGF; 5.
DR  Pfam; PF00057; ldl_recept_a; 3.
DR  PRINTS; PR00261; LDLRECEPTOR.
DR  ProDom; PD003031; Laminin_B; 2.
DR  SMART; SM00181; EGF; 6.
DR  SMART; SM00179; EGF_CA; 2.
DR  SMART; SM00180; EGF_Lam; 6.
DR  SMART; SM00409; IG; 17.
DR  SMART; SM00408; IGc2; 17.
DR  SMART; SM00281; LamB; 1.
```

```
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 3.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50835; IG_LIKE; 17.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS01209; LDLRA_1; 3.
DR   PROSITE; PS50068; LDLRA_2; 3.
KW   Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
KW   Basement membrane; Extracellular matrix; Alternative splicing;
KW   Laminin EGF-like domain.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   3375       BASEMENT MEMBRANE PROTEOGLYCAN.
FT   DOMAIN       45    130       IG-LIKE C2-TYPE 1.
FT   DOMAIN      148    184       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      189    225       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      232    269       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      271    355       IG-LIKE C2-TYPE 2.
FT   DOMAIN      384    431       LAMININ EGF-LIKE 1 (INCOMPLETE).
FT   DOMAIN      432    441       LAMININ EGF-LIKE 2 (N-TERMINAL).
FT   DOMAIN      442    633       LAMININ DOMAIN IV 1.
FT   DOMAIN      634    666       LAMININ EGF-LIKE 2 (C-TERMINAL).
FT   DOMAIN      674    720       LAMININ EGF-LIKE 3 (INCOMPLETE).
FT   DOMAIN      721    730       LAMININ EGF-LIKE 4 (N-TERMINAL).
FT   DOMAIN      731    921       LAMININ DOMAIN IV 2.
FT   DOMAIN      922    954       LAMININ EGF-LIKE 4 (C-TERMINAL).
FT   DOMAIN      955   1004       LAMININ EGF-LIKE 5.
FT   DOMAIN     1011   1060       LAMININ EGF-LIKE 6.
FT   DOMAIN     1061   1111       LAMININ EGF-LIKE 7.
FT   DOMAIN     1126   1222       IG-LIKE C2-TYPE 3.
FT   DOMAIN     1226   1311       IG-LIKE C2-TYPE 4.
FT   DOMAIN     1319   1401       IG-LIKE C2-TYPE 5.
FT   DOMAIN     1410   1499       IG-LIKE C2-TYPE 6.
FT   DOMAIN     1503   1585       IG-LIKE C2-TYPE 7.
FT   DOMAIN     1588   1680       IG-LIKE C2-TYPE 8.
FT   DOMAIN     1690   1785       IG-LIKE C2-TYPE 9.
FT   DOMAIN     1793   1878       IG-LIKE C2-TYPE 10.
FT   DOMAIN     1886   1970       IG-LIKE C2-TYPE 11.
FT   DOMAIN     1973   2069       IG-LIKE C2-TYPE 12.
FT   DOMAIN     2073   2163       IG-LIKE C2-TYPE 13.
FT   DOMAIN     2173   2260       IG-LIKE C2-TYPE 14.
FT   DOMAIN     2263   2343       IG-LIKE C2-TYPE 15.
FT   DOMAIN     2349   2435       IG-LIKE C2-TYPE 16.
FT   DOMAIN     2446   2530       IG-LIKE C2-TYPE 17.
FT   DOMAIN     2532   2713       LAMININ G-LIKE 1.
FT   DOMAIN     2793   2960       LAMININ G-LIKE 2.
FT   DOMAIN     2961   3093       GLU-RICH.
FT   DOMAIN     2972   3066       THR-RICH.
FT   DOMAIN     3180   3359       LAMININ G-LIKE 3.
FT   DISULFID     66    114       BY SIMILARITY.
FT   DISULFID    149    161       BY SIMILARITY.
FT   DISULFID    156    174       BY SIMILARITY.
FT   DISULFID    168    183       BY SIMILARITY.
FT   DISULFID    190    202       BY SIMILARITY.
FT   DISULFID    197    215       BY SIMILARITY.
FT   DISULFID    209    224       BY SIMILARITY.
FT   DISULFID    233    246       BY SIMILARITY.
FT   DISULFID    240    259       BY SIMILARITY.
FT   DISULFID    253    268       BY SIMILARITY.
FT   DISULFID    955    964       BY SIMILARITY.
FT   DISULFID    957    971       BY SIMILARITY.
FT   DISULFID    974    983       BY SIMILARITY.
FT   DISULFID    986   1002       BY SIMILARITY.
FT   DISULFID   1011   1021       BY SIMILARITY.
FT   DISULFID   1013   1027       BY SIMILARITY.
FT   DISULFID   1030   1039       BY SIMILARITY.
FT   DISULFID   1042   1058       BY SIMILARITY.
FT   DISULFID   1061   1069       BY SIMILARITY.
FT   DISULFID   1063   1079       BY SIMILARITY.
FT   DISULFID   1082   1091       BY SIMILARITY.
FT   DISULFID   1094   1109       BY SIMILARITY.
```

```
FT   DISULFID   1152   1200       BY SIMILARITY.
FT   DISULFID   1338   1384       BY SIMILARITY.
FT   DISULFID   1435   1481       BY SIMILARITY.
FT   DISULFID   1527   1573       BY SIMILARITY.
FT   DISULFID   1618   1663       BY SIMILARITY.
FT   DISULFID   1719   1767       BY SIMILARITY.
FT   DISULFID   1814   1861       BY SIMILARITY.
FT   DISULFID   1907   1954       BY SIMILARITY.
FT   DISULFID   1998   2053       BY SIMILARITY.
FT   DISULFID   2099   2147       BY SIMILARITY.
FT   DISULFID   2195   2242       BY SIMILARITY.
FT   DISULFID   2284   2329       BY SIMILARITY.
FT   DISULFID   2374   2420       BY SIMILARITY.
FT   DISULFID   2467   2514       BY SIMILARITY.
FT   CARBOHYD   1422   1422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2476   2476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2950   2950       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3143   3143       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match             22.3%;  Score 349;  DB 1;  Length 3375;
  Best Local Similarity   31.2%;  Pred. No. 6.5e-17;
  Matches   81;  Conservative   41;  Mismatches  102;  Indels   36;  Gaps    6;

Qy          30 GDLCEHEE--NPCQLREPCLHGGTCQGTR-----CLCLPGFSGPRCQQGSGHGIAESDWH 82
               ||:   :|  | |     |   | |         | |   : || |
Db        3129 GDVYSTQEPNNIC-ANSTCGMNGQCVPRNMTHYTCECKLYYDGPTCSLFK---------- 3177

Qy          83 LEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVE 142
                        | :: | |  | |:      |     |  | :  : :|   :|:||||| :
Db        3178 ---------PIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-Q 3227

Qy         143 VGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQV 202
                    | :|:||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::
Db        3228 RPTVQQMEDYISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRI 3287

Qy         203 DGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAP 262
               |    | ||| |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :
Db        3288 DNYREVDGRSTGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK---- 3343

Qy         263 PPQPLDLQHRAQAGANTRPC 282
                   |||   |  | | :||
Db        3344 ----LDLMATAIDGKNVKPC 3359


RESULT 4
AGRI_RAT
ID   AGRI_RAT       STANDARD;      PRT;  1959 AA.
AC   P25304; Q63034;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC   TISSUE=Embryonic spinal cord;
RX   MEDLINE=91222570; PubMed=1851019;
RA   Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
RT   "Structure and expression of a rat agrin.";
RL   Neuron 6:811-823(1991).
RN   [2]
RP   SEQUENCE OF 1777-1801 FROM N.A.
RX   MEDLINE=92407628; PubMed=1326608;
RA   Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
RT   "Structure and chromosomal localization of the mammalian agrin gene.";
RL   J. Neurosci. 12:3535-3544(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P25304-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P25304-2; Sequence=VSP_001365;
CC       Name=3;
CC         IsoId=P25304-3; Sequence=VSP_001366;
CC       Name=4;
CC         IsoId=P25304-4; Sequence=VSP_001367;
CC       Name=5;
CC         IsoId=P25304-5; Sequence=VSP_001368;
CC   -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
CC   -!- DEVELOPMENTAL STAGE: More abundant early in development.
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
CC       initiator.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64780; AAA40703.1; -.
DR   EMBL; M64780; AAA40702.1; ALT_INIT.
DR   EMBL; S44194; AAB23326.1; -.
DR   PIR; JH0399; AGRT.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL         1     29       POTENTIAL.
FT   CHAIN         30   1959       AGRIN.
FT   DOMAIN        65    137       KAZAL-LIKE 1.
FT   DOMAIN       141    212       KAZAL-LIKE 2.
FT   DOMAIN       213    284       KAZAL-LIKE 3.
```

```
FT   DOMAIN      287    356       KAZAL-LIKE 4.
FT   DOMAIN      361    429       KAZAL-LIKE 5.
FT   DOMAIN      430    494       KAZAL-LIKE 6.
FT   DOMAIN      495    559       KAZAL-LIKE 7.
FT   DOMAIN      563    645       KAZAL-LIKE 8.
FT   DOMAIN      688    741       LAMININ EGF-LIKE 1.
FT   DOMAIN      742    788       LAMININ EGF-LIKE 2.
FT   DOMAIN      794    864       KAZAL-LIKE 9.
FT   DOMAIN     1023   1145       SEA.
FT   DOMAIN     1220   1258       EGF-LIKE 1.
FT   DOMAIN     1263   1439       LAMININ G-LIKE 1.
FT   DOMAIN     1440   1477       EGF-LIKE 2.
FT   DOMAIN     1479   1516       EGF-LIKE 3.
FT   DOMAIN     1526   1708       LAMININ G-LIKE 2.
FT   DOMAIN     1709   1748       EGF-LIKE 4.
FT   DOMAIN     1784   1956       LAMININ G-LIKE 3.
FT   DOMAIN      869    992       SER/THR-RICH.
FT   DOMAIN     1147   1215       SER/THR-RICH.
FT   DISULFID     97    116       POTENTIAL.
FT   DISULFID    105    137       POTENTIAL.
FT   DISULFID    171    191       POTENTIAL.
FT   DISULFID    180    212       POTENTIAL.
FT   DISULFID    244    263       POTENTIAL.
FT   DISULFID    252    284       POTENTIAL.
FT   DISULFID    316    335       POTENTIAL.
FT   DISULFID    324    356       POTENTIAL.
FT   DISULFID    389    408       POTENTIAL.
FT   DISULFID    397    429       POTENTIAL.
FT   DISULFID    454    473       POTENTIAL.
FT   DISULFID    462    494       POTENTIAL.
FT   DISULFID    518    538       POTENTIAL.
FT   DISULFID    527    559       POTENTIAL.
FT   DISULFID    604    624       POTENTIAL.
FT   DISULFID    613    645       POTENTIAL.
FT   DISULFID    688    700       BY SIMILARITY.
FT   DISULFID    690    707       BY SIMILARITY.
FT   DISULFID    709    718       BY SIMILARITY.
FT   DISULFID    721    739       BY SIMILARITY.
FT   DISULFID    742    754       BY SIMILARITY.
FT   DISULFID    744    761       BY SIMILARITY.
FT   DISULFID    763    772       BY SIMILARITY.
FT   DISULFID    775    786       BY SIMILARITY.
FT   DISULFID    823    843       POTENTIAL.
FT   DISULFID    832    864       POTENTIAL.
FT   DISULFID   1224   1235       BY SIMILARITY.
FT   DISULFID   1229   1246       BY SIMILARITY.
FT   DISULFID   1248   1257       BY SIMILARITY.
FT   DISULFID   1444   1455       POTENTIAL.
FT   DISULFID   1449   1465       POTENTIAL.
FT   DISULFID   1467   1476       POTENTIAL.
FT   DISULFID   1483   1494       BY SIMILARITY.
FT   DISULFID   1488   1504       BY SIMILARITY.
FT   DISULFID   1506   1515       BY SIMILARITY.
FT   DISULFID   1713   1727       BY SIMILARITY.
FT   DISULFID   1721   1736       BY SIMILARITY.
FT   DISULFID   1738   1747       BY SIMILARITY.
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    957    957       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1144   1152       Missing (in isoform 2).
FT                                /FTId=VSP_001365.
FT   VARSPLIC   1780   1798       Missing (in isoform 3).
FT                                /FTId=VSP_001366.
FT   VARSPLIC   1788   1798       Missing (in isoform 4).
FT                                /FTId=VSP_001367.
FT   VARSPLIC   1780   1787       Missing (in isoform 5).
FT                                /FTId=VSP_001368.
FT   VARIANT     314    314       V -> VTCD (IN A VARIANT).
SQ   SEQUENCE   1959 AA;  208645 MW;  7FEFDFDAFF89CC31 CRC64;

  Query Match             21.3%;  Score 333.5;  DB 1;  Length 1959;
  Best Local Similarity   36.0%;  Pred. No. 4.5e-16;
```

```
  Matches   94;  Conservative   28;  Mismatches  108;  Indels   31;  Gaps   10;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
              |   ||  || |       | | |  |  |  |  |::|||   ||     |:     |
Db       1444 CLPNPCHGGALCQALEAGMFLCQCPPGRFGPTCADEKSPCQ-PNPCHGAAPCRVLSSGGA 1502

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
              :| |  | ||  ||             || :|       : | |:   :|   | | | |
Db       1503 KCECPLGRSGTFCQT-----------VLETAGSR----PFLADFNGFSYLELKGLHTFER 1547

Qy        115 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 174
               | |    :|:       |||||: | :    |:| ||:|| | : || | | || | |
Db       1548 DLGE-KMALEMVFLARGPSGLLLYNGQKTD--GKG-DFVSLALHNRHLEFCYDLGKGAAV 1603

Qy        175 LVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPG----PNVAVNAKGSVYIGGA 230
              : |::||  | | ||   | ||:|::||     | | ||     |:  :| |  :|||||
Db       1604 IRSKEPIALGTWVRVFLERNGRKGALQVGDGPRVLGESPKSRKVPHTMLNLKEPLYIGGA 1663

Qy        231 PDVATLT-GGRFSSGITGCVK 250
              || : |  |   ||| :| ::
Db       1664 PDFSKLARGAAVSSGFSGVIQ 1684

RESULT 5
AGRI_DISOM
ID   AGRI_DISOM     STANDARD;      PRT;  1328 AA.
AC   Q90404;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin (Fragment).
GN   AGRN.
OS   Discopyge ommata (Electric ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX   NCBI_TaxID=7785;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
RA   Snow P., McMahan U.J.;
RT   "Isolation and characterization of a cDNA that encodes an agrin
RT   homolog in the marine ray.";
RL   Mol. Cell. Neurosci. 3:406-417(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction (By similarity).
CC   -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains at least 3 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L01423; AAA49224.1; -.
DR   PIR; T43060; T43060.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
```

```
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 2.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 2.
DR   SMART; SM00280; KAZAL; 2.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT   NON_TER       1      1
FT   DOMAIN       79    132       LAMININ EGF-LIKE 1.
FT   DOMAIN      133    179       LAMININ EGF-LIKE 2.
FT   DOMAIN      411    533       SEA.
FT   DOMAIN      608    644       EGF-LIKE 1.
FT   DOMAIN      649    825       LAMININ G-LIKE 1.
FT   DOMAIN      865    902       EGF-LIKE 2.
FT   DOMAIN      914   1096       LAMININ G-LIKE 2.
FT   DOMAIN     1097   1135       EGF-LIKE 3.
FT   DOMAIN     1146   1324       LAMININ G-LIKE 3.
FT   DISULFID     79     91       BY SIMILARITY.
FT   DISULFID     81     98       BY SIMILARITY.
FT   DISULFID    100    109       BY SIMILARITY.
FT   DISULFID    112    130       BY SIMILARITY.
FT   DISULFID    133    145       BY SIMILARITY.
FT   DISULFID    135    152       BY SIMILARITY.
FT   DISULFID    154    163       BY SIMILARITY.
FT   DISULFID    166    177       BY SIMILARITY.
FT   DISULFID    612    623       BY SIMILARITY.
FT   DISULFID    617    632       BY SIMILARITY.
FT   DISULFID    634    643       BY SIMILARITY.
FT   DISULFID    869    880       BY SIMILARITY.
FT   DISULFID    874    890       BY SIMILARITY.
FT   DISULFID    892    901       BY SIMILARITY.
FT   DISULFID   1101   1114       BY SIMILARITY.
FT   DISULFID   1108   1123       BY SIMILARITY.
FT   DISULFID   1125   1134       BY SIMILARITY.
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1328 AA;  144018 MW;  79D81C1AF2A71C18 CRC64;

  Query Match             20.1%;  Score 314.5;  DB 1;  Length 1328;
  Best Local Similarity   31.4%;  Pred. No. 6.7e-15;
  Matches   85;  Conservative   44;  Mismatches  105;  Indels   37;  Gaps   12;

Qy            1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
                |   ||::   |       | |     | |  |  : |||    ||     |      |:
Db          830 CSPNPCKNRGKCHMKEAEMFHCESVGEFSGPTCADKHNPCD-PNPCHQSANCMVLPEGGS 888

Qy           56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
                :| |  |  |  |::     ::|::         |    :   |:   :|   | | |
Db          889 KCECPMGREGELCER-----VSEAE--------QDQGKAFIPEFNGLSYLEMNGIHTF-- 933

Qy          115 SLPEVPETIELEV--RTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGE 172
                 : :: : : :||       :|:: : | :    |:| ||:|| |:||:| |:| || |
Db          934 -VSDLLQKLSMEVIFLAKDPNGMIFYNGQKTD--GRG-DFVSLNLRDGYLEFKYDLGKGA 989

Qy          173 ARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP----GPNVAVNAKGSVYIG 228
                | | |: ||    |: ||  | ||:| :::: :||||| ||     |: |:| | : |:|
Db          990 AVLRSKAPIPLNVWNVVTVERNGRKGLMKINKDELVSGESPKSRKAPHTALNLKEAFYVG 1049

Qy          229 GAPD---VATLTGGRFSSGITGCVKNLVLHS 256
                ||||    |   |    || ||  :: | | |
Db         1050 GAPDFNKFARAAG--IISGFTGAIQKLSLKS 1078

RESULT 6
AGRI_CHICK
ID   AGRI_CHICK     STANDARD;      PRT;  1955 AA.
AC   P31696;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Agrin precursor.
GN   AGRN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92232297; PubMed=1314620;
RA   Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahan U.J.;
RT   "cDNA that encodes active agrin.";
RL   Neuron 8:677-689(1992).
RN   [2]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=92232298; PubMed=1314621;
RA   Ruegg M.A., Tsim K.W.K., Horton S.E., Kroeger S., Escher G.,
RA   Gensch E.M., McMahan U.J.;
RT   "The agrin gene codes for a family of basal lamina proteins that
RT   differ in function and distribution.";
RL   Neuron 8:691-699(1992).
CC   -!- FUNCTION: Component of the basal lamina that causes the
CC       aggregation of acetylcholine receptors and acetylcholine-esterase
CC       on the surface of muscle fibers of the neuromuscular junction.
CC   -!- SUBUNIT: Binds to laminin.
CC   -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC       junction.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Additional isoforms seem to exist. Isoforms differ in
CC         their acetylcholine receptor clustering activity;
CC       Name=1;
CC         IsoId=P31696-1; Sequence=Displayed;
CC       Name=2; Synonyms=Agrin-related protein 1;
CC         IsoId=P31696-2; Sequence=VSP_001370;
CC       Name=3; Synonyms=Agrin-related protein 2;
CC         IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
CC   -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC       linked oligosaccharides (By similarity).
CC   -!- SIMILARITY: Contains 9 Kazal-like domains.
CC   -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M94271; AAA48585.1; -.
DR   EMBL; M97371; AAA48586.1; -.
DR   EMBL; M97372; -; NOT_ANNOTATED_CDS.
DR   PIR; JH0591; AGCH.
DR   HSSP; P00740; 1EDM.
```

```
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 8.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW   Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1955       AGRIN.
FT   DOMAIN       54    126       KAZAL-LIKE 1.
FT   DOMAIN      130    201       KAZAL-LIKE 2.
FT   DOMAIN      202    273       KAZAL-LIKE 3.
FT   DOMAIN      276    344       KAZAL-LIKE 4.
FT   DOMAIN      350    418       KAZAL-LIKE 5.
FT   DOMAIN      419    483       KAZAL-LIKE 6.
FT   DOMAIN      484    548       KAZAL-LIKE 7.
FT   DOMAIN      551    633       KAZAL-LIKE 8.
FT   DOMAIN      675    728       LAMININ EGF-LIKE 1.
FT   DOMAIN      729    775       LAMININ EGF-LIKE 2.
FT   DOMAIN      781    851       KAZAL-LIKE 9.
FT   DOMAIN     1026   1148       SEA.
FT   DOMAIN     1229   1265       EGF-LIKE 1.
FT   DOMAIN     1270   1445       LAMININ G-LIKE 1.
FT   DOMAIN     1446   1483       EGF-LIKE 2.
FT   DOMAIN     1485   1522       EGF-LIKE 3.
FT   DOMAIN     1532   1713       LAMININ G-LIKE 2.
FT   DOMAIN     1714   1752       EGF-LIKE 4.
FT   DOMAIN     1776   1952       LAMININ G-LIKE 3.
FT   DOMAIN      856    995       SER/THR-RICH.
FT   DOMAIN     1150   1219       SER/THR-RICH.
FT   DISULFID     86    105       POTENTIAL.
FT   DISULFID     94    126       POTENTIAL.
FT   DISULFID    160    180       POTENTIAL.
FT   DISULFID    169    201       POTENTIAL.
FT   DISULFID    233    252       POTENTIAL.
FT   DISULFID    241    273       POTENTIAL.
FT   DISULFID    304    323       POTENTIAL.
FT   DISULFID    312    344       POTENTIAL.
FT   DISULFID    378    397       POTENTIAL.
FT   DISULFID    386    418       POTENTIAL.
FT   DISULFID    443    462       POTENTIAL.
FT   DISULFID    451    483       POTENTIAL.
FT   DISULFID    507    527       POTENTIAL.
FT   DISULFID    516    548       POTENTIAL.
FT   DISULFID    592    612       POTENTIAL.
FT   DISULFID    601    633       POTENTIAL.
FT   DISULFID    675    687       BY SIMILARITY.
FT   DISULFID    677    694       BY SIMILARITY.
```

```
FT   DISULFID    696    705       BY SIMILARITY.
FT   DISULFID    708    726       BY SIMILARITY.
FT   DISULFID    729    741       BY SIMILARITY.
FT   DISULFID    731    748       BY SIMILARITY.
FT   DISULFID    750    759       BY SIMILARITY.
FT   DISULFID    762    773       BY SIMILARITY.
FT   DISULFID    810    830       POTENTIAL.
FT   DISULFID    819    851       POTENTIAL.
FT   DISULFID   1233   1244       BY SIMILARITY.
FT   DISULFID   1238   1253       BY SIMILARITY.
FT   DISULFID   1255   1264       BY SIMILARITY.
FT   DISULFID   1450   1461       BY SIMILARITY.
FT   DISULFID   1455   1471       BY SIMILARITY.
FT   DISULFID   1473   1482       BY SIMILARITY.
FT   DISULFID   1489   1500       BY SIMILARITY.
FT   DISULFID   1494   1510       BY SIMILARITY.
FT   DISULFID   1512   1521       BY SIMILARITY.
FT   DISULFID   1718   1731       BY SIMILARITY.
FT   DISULFID   1725   1740       BY SIMILARITY.
FT   DISULFID   1742   1751       BY SIMILARITY.
FT   CARBOHYD    390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1648   1651       Missing (in isoform 3).
FT                                /FTId=VSP_001369.
FT   VARSPLIC   1783   1793       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_001370.
FT   CONFLICT   1129   1131       RTI -> SIL (IN REF. 1; AAA48586).
SQ   SEQUENCE   1955 AA;  211411 MW;  B4DEB27C23422581 CRC64;

  Query Match             19.3%;  Score 301.5;  DB 1;  Length 1955;
  Best Local Similarity   27.4%;  Pred. No. 8.6e-14;
  Matches   87;  Conservative   38;  Mismatches  108;  Indels   85;  Gaps   13;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTC-----QGT 55
              |   || |||:|       | | |   : |  |  | |||    ||    ||      |
Db       1450 CHPNPCHHGASCHVKEAEMFHCECLHSYTGPTCADERNPCD-PTPCHISATCLVLPEGGA 1508

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRS 115
               | |  |  |  |::     : | |                                  :
Db       1509 MCACPMGREGEFCER-----VTEQD---------------------------------HT 1530

Qy        116 LPEVPE----------------------TIELEVRTSTASGLLLWQGVEVGEAGQGKDFI 153
              :| :||                      ::|:     : ||:: : | :    |:| ||:
Db       1531 MPFLPEFNGFSYLELNGLQTLFLTCRQMSMEVVFLAKSPSGMIFYNGQKTD--GKG-DFV 1587

Qy        154 SLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSP 213
              || | ||:| :|| || | | | |::|:    |  |   | ||:| ::::  | | | ||
Db       1588 SLALHDGYLEYRYDLGKGAAVLRSKEPVPLNTWISVLLERSGRKGVMRINNGERVMGESP 1647

Qy        214 G----PNVAVNAKGSVYIGGAPDVATLT-GGRFSSGITGCVKNLVLHSARPGAPPPQPLD 268
                   |:  :| |   |:||||| : |      |:   | |: : :     | |    |
Db       1648 KSRKVPHAFLNLKEPFYVGGAPDFSKLARAAAISTSFYGAVQRISI----KGVP---LLK 1700

Qy        269 LQH-RAQAGANT---RPC 282
               || |:    :|    ||
Db       1701 EQHIRSAVEISTFRAHPC 1718


RESULT 7
FAT_DROME
ID   FAT_DROME      STANDARD;      PRT;  5147 AA.
AC   P33450; Q9VQX5;
DT   01-FEB-1994 (Rel. 28, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cadherin-related tumor suppressor precursor (Fat protein).
GN   FT OR CG3352.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
```

```
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92069752; PubMed=1959133;
RA   Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J.,
RA   Goodman C.S.;
RT   "The fat tumor suppressor gene in Drosophila encodes a novel member
RT   of the cadherin gene superfamily.";
RL   Cell 67:853-868(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: Could function as a cell-adhesion protein. Acts as a
CC       tumor suppressor. Required for correct morphogenesis.
CC   -!- SIMILARITY: Contains 34 cadherin domains.
CC   -!- SIMILARITY: Contains 5 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M80537; AAA28530.1; -.
DR   EMBL; AE003577; AAF51036.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0001075; ft.
DR   GO; GO:0005887; C:integral to plasma membrane; NAS.
DR   GO; GO:0008014; F:calcium-dependent cell adhesion molecule ac. . .; NAS.
DR   GO; GO:0016339; P:calcium-dependent cell-cell adhesion; NAS.
DR   GO; GO:0008283; P:cell proliferation; IMP.
```

```
DR   GO; GO:0000904; P:cellular morphogenesis during differentiation; IMP.
DR   GO; GO:0045317; P:equator specification; IMP.
DR   GO; GO:0045198; P:establishment of epithelial cell polarity; IMP.
DR   GO; GO:0007446; P:imaginal disc growth; IMP.
DR   GO; GO:0018149; P:protein-protein cross-linking; IPI.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00028; cadherin; 34.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00054; laminin_G; 2.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 34.
DR   SMART; SM00181; EGF; 4.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00232; CADHERIN_1; 22.
DR   PROSITE; PS50268; CADHERIN_2; 34.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50026; EGF_3; 4.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Cell adhesion; Signal; Transmembrane; Glycoprotein; Calcium-binding;
KW   Repeat; EGF-like domain.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36   5147       CADHERIN-RELATED TUMOR SUPPRESSOR.
FT   DOMAIN       36   4583       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   4584   4609       POTENTIAL.
FT   DOMAIN     4610   5147       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       36    156       CADHERIN 1.
FT   DOMAIN      157    270       CADHERIN 2.
FT   DOMAIN      271    382       CADHERIN 3.
FT   DOMAIN      383    494       CADHERIN 4.
FT   DOMAIN      495    599       CADHERIN 5.
FT   DOMAIN      600    708       CADHERIN 6.
FT   DOMAIN      709    820       CADHERIN 7.
FT   DOMAIN      821    942       CADHERIN 8.
FT   DOMAIN      943   1049       CADHERIN 9.
FT   DOMAIN     1050   1153       CADHERIN 10.
FT   DOMAIN     1154   1278       CADHERIN 11.
FT   DOMAIN     1279   1384       CADHERIN 12.
FT   DOMAIN     1385   1489       CADHERIN 13.
FT   DOMAIN     1490   1601       CADHERIN 14.
FT   DOMAIN     1602   1713       CADHERIN 15.
FT   DOMAIN     1714   1823       CADHERIN 16.
FT   DOMAIN     1824   1922       CADHERIN 17.
FT   DOMAIN     1923   2027       CADHERIN 18.
FT   DOMAIN     2028   2167       CADHERIN 19.
FT   DOMAIN     2168   2278       CADHERIN 20.
FT   DOMAIN     2279   2385       CADHERIN 21.
FT   DOMAIN     2386   2491       CADHERIN 22.
FT   DOMAIN     2492   2596       CADHERIN 23.
FT   DOMAIN     2597   2703       CADHERIN 24.
FT   DOMAIN     2704   2810       CADHERIN 25.
FT   DOMAIN     2811   2913       CADHERIN 26.
FT   DOMAIN     2914   3013       CADHERIN 27.
FT   DOMAIN     3014   3124       CADHERIN 28.
FT   DOMAIN     3125   3229       CADHERIN 29.
FT   DOMAIN     3230   3334       CADHERIN 30.
FT   DOMAIN     3335   3439       CADHERIN 31.
FT   DOMAIN     3440   3545       CADHERIN 32.
FT   DOMAIN     3546   3651       CADHERIN 33.
FT   DOMAIN     3652   3756       CADHERIN 34.
FT   DOMAIN     3950   4011       EGF-LIKE 1.
FT   DOMAIN     4013   4049       EGF-LIKE 2.
FT   DOMAIN     4052   4090       EGF-LIKE 3.
FT   DOMAIN     4092   4128       EGF-LIKE 4.
FT   DOMAIN     4129   4320       LAMININ G-LIKE 1.
FT   DOMAIN     4321   4362       EGF-LIKE 5.
FT   DOMAIN     4402   4569       LAMININ G-LIKE 2.
FT   DISULFID   3954   3966       BY SIMILARITY.
```

```
FT   DISULFID    3960   3999       BY SIMILARITY.
FT   DISULFID    4001   4010       BY SIMILARITY.
FT   DISULFID    4017   4028       BY SIMILARITY.
FT   DISULFID    4022   4037       BY SIMILARITY.
FT   DISULFID    4039   4048       BY SIMILARITY.
FT   DISULFID    4056   4067       BY SIMILARITY.
FT   DISULFID    4061   4078       BY SIMILARITY.
FT   DISULFID    4080   4089       BY SIMILARITY.
FT   DISULFID    4096   4107       BY SIMILARITY.
FT   DISULFID    4101   4116       BY SIMILARITY.
FT   DISULFID    4118   4127       BY SIMILARITY.
FT   DISULFID    4325   4341       BY SIMILARITY.
FT   DISULFID    4334   4350       BY SIMILARITY.
FT   DISULFID    4352   4361       BY SIMILARITY.
FT   CARBOHYD     239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     257    257       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     276    276       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     461    461       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     605    605       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1155   1155       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1367   1367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1458   1458       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1751   1751       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1831   1831       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1880   1880       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2080   2080       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2171   2171       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2247   2247       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2290   2290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2437   2437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2581   2581       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2799   2799       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2920   2920       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2946   2946       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2967   2967       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    3167   3167       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    3303   3303       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    3386   3386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    3389   3389       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    3525   3525       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match            17.2%;  Score 269;  DB 1;  Length 5147;
  Best Local Similarity   29.7%;  Pred. No. 5.1e-11;
  Matches   94;  Conservative   39;  Mismatches  102;  Indels   82;  Gaps   19;

Qy          1 CERQPCQHGATCM--PAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTC----QG 54
              |  :||::| :|   | |   | |||| ||:|: ||   : |:   |||||| |     |
Db       4056 CYSKPCRNGGSCQRSPDGSSYF-CLCRPGFRGNQCESVSDSCR-PNPCLHGGLCVSLKPG 4113

Qy         55 TRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSR 114
               :| | ||  |  |:: |                      ||  |     ::||
Db       4114 YKCNCTPGRYGRHCERFS----------------------YG--FQPLSYMTFP------ 4143

Qy        115 SLPEVPETIELEVRTSTASGLLLWQ-GVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEA 173
              :|      | :   |:  : |||:  |:   ::|   ||::: |  |   |     || |
Db       4144 ALDVTTNDISIVFATTKPNSLLLYNYGM---QSGGRSDFLAIELVHGRAYF----SSGGA 4196

Qy        174 R-----LVSEDPINDGEWHRVTALREGRRGSIQV-----DGE---ELVSGRSP------G 214
              |     :::   : || ||:||| | ||  |: |      |:   | : | |       |
Db       4197 RTAISTVIAGRNLADGGWHKVTATRNGRVMSLSVAKCADSGDVCTECLPGDSSCYADEVG 4256

Qy        215 PNVAVN-AKGSVYIGG---APDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQPLDL- 269
              |   :|  |  : |||   |  :    |   |  : ||     |||   |    : |:|
Db       4257 PVGTLNFNKQPLMIGGLSSADPILERPGQVHSDDLVGC-----LHSVHIGG---RALNLS 4308

Qy        270 ----QHRAQAGANTRPC 282
                  |    || | : |
Db       4309 LPLQQKGILAGCNRQAC 4325
```

```
RESULT 8
SLT2_HUMAN
ID   SLT2_HUMAN     STANDARD;      PRT;  1529 AA.
AC   O94813; O95710; Q9Y5Q7;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Slit homolog 2 protein precursor (h-Slit-2).
GN   SLIT2 OR SLIL2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC   TISSUE=Fetal lung;
RX   MEDLINE=99033071; PubMed=9813312;
RA   Itoh A., Miyabayashi T., Ohno M., Sakano S.;
RT   "Cloning and expressions of three mammalian homologues of Drosophila
RT   slit suggest possible roles for slit in the formation and maintenance
RT   of the nervous system.";
RL   Brain Res. Mol. Brain Res. 62:175-186(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 3), AND TISSUE SPECIFICITY.
RC   TISSUE=Fetal brain, and Fetal kidney;
RX   MEDLINE=99279238; PubMed=10349621;
RA   Holmes G.P., Negus K., Burridge L., Raman S., Algar E., Yamada T.,
RA   Little M.H.;
RT   "Distinct but overlapping expression patterns of two vertebrate slit
RT   homologs implies functional roles in CNS development and
RT   organogenesis.";
RL   Mech. Dev. 79:57-72(1998).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 1122-1129, FUNCTION,
RP   SUBUNIT, AND SUBCELLULAR LOCATION.
RC   TISSUE=Fetal brain;
RX   MEDLINE=99200391; PubMed=10102268;
RA   Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA   Tessier-Lavigne M., Kidd T.;
RT   "Slit proteins bind Robo receptors and have an evolutionarily
RT   conserved role in repulsive axon guidance.";
RL   Cell 96:795-806(1999).
CC   -!- FUNCTION: Plays a role in axon guidance as a repulsive ligand for
CC       Robo receptors preventing inappropriate midline crossing.
CC   -!- SUBUNIT: Binds robo proteins with high affinity.
CC   -!- SUBCELLULAR LOCATION: Secreted. The C-terminal cleavage protein is
CC       more diffusible than the larger N-terminal protein that is more
CC       tightly cell associated.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=O94813-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O94813-2; Sequence=VSP_050035, VSP_050036;
CC       Name=3;
CC         IsoId=O94813-3; Sequence=VSP_050036;
CC   -!- TISSUE SPECIFICITY: Fetal lung and kidney, and adult spinal cord.
CC       Weak expression in adult adrenal gland, thyroid, trachea and other
CC       tissues examined.
CC   -!- SIMILARITY: Contains 7 EGF-like domains.
CC   -!- SIMILARITY: Contains 23 leucine-rich (LRR) repeats.
CC   -!- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.
CC   -!- SIMILARITY: Contains 1 laminin G-like domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB017168; BAA35185.1; -.
```

```
DR   EMBL; AF055585; AAD04309.1; -.
DR   EMBL; AF133270; AAD25539.1; -.
DR   HSSP; P00743; 1CCF.
DR   Genew; HGNC:11086; SLIT2.
DR   MIM; 603746; -.
DR   GO; GO:0005576; C:extracellular; IEP.
DR   GO; GO:0045499; F:chemorepellant activity; IMP.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0008347; P:glia cell migration; IMP.
DR   GO; GO:0007509; P:mesoderm migration; IMP.
DR   GO; GO:0008045; P:motor axon guidance; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   GO; GO:0008038; P:neuronal cell recognition; TAS.
DR   GO; GO:0007608; P:olfaction; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00008; EGF; 8.
DR   Pfam; PF00054; laminin_G; 1.
DR   Pfam; PF00560; LRR; 18.
DR   Pfam; PF01463; LRRCT; 4.
DR   Pfam; PF01462; LRRNT; 4.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00274; FOLN; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00369; LRR_TYP; 8.
DR   SMART; SM00082; LRRCT; 4.
DR   SMART; SM00013; LRRNT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS50026; EGF_3; 9.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Neurogenesis; Glycoprotein; Signal; Alternative splicing;
KW   EGF-like domain; Repeat; Leucine-rich repeat.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1529       SLIT PROTEIN.
FT   CHAIN        31   1121       SLIT PROTEIN N-PRODUCT.
FT   CHAIN      1122   1529       SLIT PROTEIN C-PRODUCT.
FT   SITE       1121   1122       CLEAVAGE.
FT   REPEAT       54     77       LRR 1.
FT   REPEAT       78    101       LRR 2.
FT   REPEAT      102    125       LRR 3.
FT   REPEAT      126    149       LRR 4.
FT   REPEAT      151    173       LRR 5.
FT   REPEAT      174    197       LRR 6.
FT   REPEAT      199    226       LRR 7.
FT   REPEAT      299    322       LRR 8.
FT   REPEAT      323    346       LRR 9.
FT   REPEAT      347    370       LRR 10.
FT   REPEAT      372    394       LRR 11.
FT   REPEAT      395    418       LRR 12.
FT   REPEAT      511    531       LRR 13.
FT   REPEAT      532    556       LRR 14.
FT   REPEAT      557    580       LRR 15.
FT   REPEAT      582    604       LRR 16.
FT   REPEAT      605    628       LRR 17.
FT   REPEAT      629    652       LRR 18.
FT   REPEAT      753    776       LRR 19.
FT   REPEAT      777    799       LRR 20.
FT   REPEAT      800    823       LRR 21.
FT   REPEAT      824    847       LRR 22.
FT   REPEAT      849    872       LRR 23.
FT   DOMAIN      918    955       EGF-LIKE 1.
FT   DOMAIN      957    996       EGF-LIKE 2.
FT   DOMAIN      998   1034       EGF-LIKE 3, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN     1036   1074       EGF-LIKE 4.
FT   DOMAIN     1076   1112       EGF-LIKE 5, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN     1121   1157       EGF-LIKE 6.
FT   DOMAIN     1160   1333       LAMININ G-LIKE.
FT   DOMAIN     1332   1368       EGF-LIKE 7.
FT   DOMAIN     1453   1528       CTCK.
FT   DISULFID    922    933       BY SIMILARITY.
FT   DISULFID    927    943       BY SIMILARITY.
FT   DISULFID    945    954       BY SIMILARITY.
FT   DISULFID    961    972       BY SIMILARITY.
FT   DISULFID    966    984       BY SIMILARITY.
FT   DISULFID    986    995       BY SIMILARITY.
FT   DISULFID   1002   1013       BY SIMILARITY.
FT   DISULFID   1007   1022       BY SIMILARITY.
FT   DISULFID   1024   1033       BY SIMILARITY.
FT   DISULFID   1040   1053       BY SIMILARITY.
FT   DISULFID   1047   1062       BY SIMILARITY.
FT   DISULFID   1064   1073       BY SIMILARITY.
FT   DISULFID   1080   1091       BY SIMILARITY.
FT   DISULFID   1085   1100       BY SIMILARITY.
FT   DISULFID   1102   1111       BY SIMILARITY.
FT   DISULFID   1125   1136       BY SIMILARITY.
FT   DISULFID   1130   1145       BY SIMILARITY.
FT   DISULFID   1147   1156       BY SIMILARITY.
FT   DISULFID   1336   1346       BY SIMILARITY.
FT   DISULFID   1341   1356       BY SIMILARITY.
FT   DISULFID   1358   1367       BY SIMILARITY.
FT   CARBOHYD     66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    186    186       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    794    794       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    799    799       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1009   1009       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1010   1010       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1019   1019       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1183   1183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1266   1266       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1300   1300       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    258    258       S -> SDEEE (in isoform 2).
FT                                /FTId=VSP_050035.
FT   VARSPLIC    480    487       Missing (in isoform 2 and isoform
FT                                3).
FT                                /FTId=VSP_050036.
FT   CONFLICT    226    226       Q -> K (IN REF. 2).
FT   CONFLICT    607    610       SLKT -> KPQN (IN REF. 3).
FT   CONFLICT    634    634       L -> M (IN REF. 3).
SQ   SEQUENCE   1529 AA;  169869 MW;  5D19CC5E7FD461BA CRC64;

  Query Match             16.2%;  Score 253.5;  DB 1;  Length 1529;
  Best Local Similarity   24.8%;  Pred. No. 1.7e-10;
  Matches   92;  Conservative   48;  Mismatches  100;  Indels  131;  Gaps   18;

Qy          1 CERQPCQHGATC-MPAGEYE-FQCLCRDGFKGDLCE-----HEENPC------------- 40
              |   ||:|| || :  || : | |:| |||:|: ||      |:| |
Db        961 CISNPCKHGGTCHLKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNY 1020

Qy         41 ---------------------QLREPCLHGGTC----QGTRCLCLPGFSGP--------- 66
                                   |   || |   |    :| :| | ||: |
Db       1021 TCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDC 1080

Qy         67 ---RCQQGSGHGIAESDWHL---EGSGG-----------------------NDA------ 91
                 :|: |:    | : :     ||  |                       | |
Db       1081 QDNKCKNGAHCTDAVNGYTCICPEGYSGLFCEFSPPMVLPRTSPCDNFDCQNGAQCIVRI 1140
```

```
Qy        92 --------PGQYGA---------YFHDDGFLAFPGHVFSRSLPEVPET-IELEVRTSTAS 133
                     ||  |          : : : :|  |      |    |:| | |:: |   |
Db      1141 NEPICQCLPGYQGEKCEKLVSVNFINKESYLQIP------SAKVRPQTNITLQIATDEDS 1194

Qy       134 GLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLV-SEDPINDGEWHRVTAL 192
             |:||::|         || |:: |  | :   |  ||  |  : | : |||| :| |  |
Db      1195 GILLYKG--------DKDHIAVELYRGRVRASYDTGSHPASAIYSVETINDGNFHIVELL 1246

Qy       193 REGRRGSIQVDG--EELVSGRSPGPNVAVNAKGSVYIGGAP---DVATL--TGGRFSSGI 245
                :  |: |||   ::::  |      :|    :|:|| |   :||:|    |:  :
Db      1247 ALDQSLSLSVDGGNPKIITNLS--KQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSF 1304

Qy       246 TGCVKNLVLHS 256
              ||::|| ::|
Db      1305 HGCIRNLYINS 1315

RESULT 9
FAT2_DROME
ID   FAT2_DROME     STANDARD;      PRT;  4705 AA.
AC   Q9VW71; Q95S51;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative fat-like cadherin-related tumor suppressor homolog
DE   precursor.
GN   FAT2 OR CG7749.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   REVISIONS.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [3]
RP   SEQUENCE OF 3837-4705 FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Ovary;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- SIMILARITY: Contains 34 cadherin domains.
CC   -!- SIMILARITY: Contains 5 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 laminin G-like domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE003515; AAF49078.2; -.
DR   EMBL; AY060955; AAL28503.1; ALT_INIT.
DR   EMBL; AY118666; AAM50035.1; ALT_INIT.
DR   HSSP; P15116; 1NCI.
DR   FlyBase; FBgn0036930; fat2.
DR   GO; GO:0005887; C:integral to plasma membrane; ISS.
DR   GO; GO:0008014; F:calcium-dependent cell adhesion molecule ac. . .; ISS.
DR   GO; GO:0016339; P:calcium-dependent cell-cell adhesion; ISS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00028; cadherin; 31.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 34.
DR   SMART; SM00181; EGF; 6.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00232; CADHERIN_1; 18.
DR   PROSITE; PS50268; CADHERIN_2; 34.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50026; EGF_3; 5.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Hypothetical protein; Cell adhesion; Signal; Glycoprotein;
KW   Transmembrane; Calcium; Calcium-binding; Repeat; EGF-like domain.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36   4705       PUTATIVE FAT-LIKE CADHERIN-RELATED TUMOR
FT                                SUPPRESSOR HOMOLOG.
FT   DOMAIN       36   1647       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1648   1668       POTENTIAL.
```

```
FT   DOMAIN      1669   4705       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        60    180       CADHERIN 1.
FT   DOMAIN       181    288       CADHERIN 2.
FT   DOMAIN       285    397       CADHERIN 3.
FT   DOMAIN       398    504       CADHERIN 4.
FT   DOMAIN       505    610       CADHERIN 5.
FT   DOMAIN       611    713       CADHERIN 6.
FT   DOMAIN       770    874       CADHERIN 7.
FT   DOMAIN       875    977       CADHERIN 8.
FT   DOMAIN       978   1085       CADHERIN 9.
FT   DOMAIN      1086   1195       CADHERIN 10.
FT   DOMAIN      1191   1296       CADHERIN 11.
FT   DOMAIN      1297   1402       CADHERIN 12.
FT   DOMAIN      1405   1503       CADHERIN 13.
FT   DOMAIN      1504   1609       CADHERIN 14.
FT   DOMAIN      1610   1714       CADHERIN 15.
FT   DOMAIN      1715   1812       CADHERIN 16.
FT   DOMAIN      1813   1929       CADHERIN 17.
FT   DOMAIN      1930   2030       CADHERIN 18.
FT   DOMAIN      2031   2137       CADHERIN 19.
FT   DOMAIN      2138   2238       CADHERIN 20.
FT   DOMAIN      2239   2338       CADHERIN 21.
FT   DOMAIN      2339   2465       CADHERIN 22.
FT   DOMAIN      2466   2567       CADHERIN 23.
FT   DOMAIN      2568   2670       CADHERIN 24.
FT   DOMAIN      2671   2779       CADHERIN 25.
FT   DOMAIN      2780   2876       CADHERIN 26.
FT   DOMAIN      2877   2983       CADHERIN 27.
FT   DOMAIN      2984   3088       CADHERIN 28.
FT   DOMAIN      3084   3185       CADHERIN 29.
FT   DOMAIN      3186   3289       CADHERIN 30.
FT   DOMAIN      3290   3394       CADHERIN 31.
FT   DOMAIN      3395   3499       CADHERIN 32.
FT   DOMAIN      3500   3604       CADHERIN 33.
FT   DOMAIN      3605   3712       CADHERIN 34.
FT   DOMAIN      3819   3879       EGF-LIKE 1.
FT   DOMAIN      3881   3919       EGF-LIKE 2.
FT   DOMAIN      3937   4121       LAMININ G-LIKE.
FT   DOMAIN      4129   4166       EGF-LIKE 3.
FT   DOMAIN      4168   4205       EGF-LIKE 4.
FT   DOMAIN      4243   4279       EGF-LIKE 5.
FT   DISULFID    3823   3835       POTENTIAL.
FT   DISULFID    3830   3867       POTENTIAL.
FT   DISULFID    3869   3878       POTENTIAL.
FT   DISULFID    3885   3896       POTENTIAL.
FT   DISULFID    3890   3907       POTENTIAL.
FT   DISULFID    3909   3918       POTENTIAL.
FT   DISULFID    4133   4144       POTENTIAL.
FT   DISULFID    4138   4154       POTENTIAL.
FT   DISULFID    4156   4165       POTENTIAL.
FT   DISULFID    4172   4183       POTENTIAL.
FT   DISULFID    4177   4193       POTENTIAL.
FT   DISULFID    4195   4204       POTENTIAL.
FT   DISULFID    4247   4258       POTENTIAL.
FT   DISULFID    4252   4267       POTENTIAL.
FT   DISULFID    4269   4278       POTENTIAL.
FT   CARBOHYD      65     65       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     156    156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     364    364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     779    779       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     843    843       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     923    923       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1106   1106       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1198   1198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1312   1312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1439   1439       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1473   1473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1511   1511       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    3962   3962       G -> E (IN REF. 3; AAL28503).
SQ   SEQUENCE    4705 AA;  524564 MW;  6D387A489D2C33DE CRC64;

  Query Match             15.5%;  Score 242;  DB 1;  Length 4705;
  Best Local Similarity   27.0%;  Pred. No. 3.9e-09;
```

```
  Matches   69;  Conservative   43;  Mismatches   88;  Indels   56;  Gaps   10;

Qy          22 CLCRDGFKGDLCEHEEN-----PCQLREPCLHGGTCQGTRCLCLPGFSGPRCQQGSGHGI 76
               |:|| || |  |:   |     ||  :  |:  |:  | :|:|  ||||  |:: |
Db        3867 CVCRPGFDGKHCKETVNACSTDPCSPQRICMPSGSALGYQCVCPKGFSGTYCERKSSKCS 3926

Qy          77 AES-DWHL---EGSGGNDAPGQYGAYFH------------DDGFLAFPGHVFSRSLPEVP 120
                || |  |      ||        :| |            ::||
Db        3927 NESCDMGLFTAVSFGGK-------SYAHYKINKVKAKFTLENGF---------------- 3963

Qy         121 ETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARL-VSED 179
                :  |::||   :| ||:   :|       |:  | : :| : :|: |||||  : ||
Db        3964 -SYSLQIRTVQQTGTLLYASGKV-------DYNILEIINGAVQYRFDLGSGEGVISVSSI 4015

Qy         180 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGS-VYIGG--APDVATL 236
                |:|||||:::  |      : || : :  | :|| |  :| : : :::|    |  : :
Db        4016 NISDGEWHQISLERSLNSAKVMVDNKHVSHGSAPGVNGILNIQSNDIFVGAEVRPHPSII 4075

Qy         237 TGGRFSSGITGCVKNL 252
                      |  ||: |:
Db        4076 GYEDIQRGFIGCMANI 4091


RESULT 10
STAN_DROME
ID   STAN_DROME     STANDARD;      PRT;  3579 AA.
AC   Q9V5N8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protocadherin-like wing polarity protein stan precursor (Starry night
DE   protein) (Flamingo protein).
GN   STAN OR FMI OR CG11895.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP   STAGE.
RC   TISSUE=Embryo;
RX   MEDLINE=20025940; PubMed=10556066;
RA   Chae J.W., Kim M.-J., Goo J.H., Collier S., Gubb D., Charlton J.,
RA   Adler P.N., Park W.J.;
RT   "The Drosophila tissue polarity gene starry night encodes a member of
RT   the protocadherin family.";
RL   Development 126:5421-5429(1999).
RN   [2]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=99418630; PubMed=10490098;
RA   Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA   Takeichi M., Uemura T.;
RT   "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT   polarity under the control of frizzled.";
RL   Cell 98:585-595(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
```

```
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   REVISIONS.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC   -!- FUNCTION: Involved in the fz signaling pathway that controls wing
CC       tissue polarity. Also mediates homophilic cell adhesion. May play
CC       a role in initiating prehair morphogenesis. May play a critical
CC       role in tissue polarity and in formation of normal dendrite
CC       fields.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: In the pupal wing, expressed at relatively
CC       even levels in all regions. Abundant in 6-9 hour embryos.
CC       Expressed at higher levels in pupae than larvae.
CC   -!- DEVELOPMENTAL STAGE: At 12 hours after puparium formation (apf),
CC       expressed evenly at cell boundaries. By 30 hours apf, expression
CC       is concentrated at proximal and distal cell boundaries with little
CC       or no expression at anterior and posterior boundaries. When
CC       prehairs emerge at 30-36 hours apf, expression becomes evenly
CC       distributed again along the whole cell boundary.
CC   -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC   -!- SIMILARITY: Contains 8 cadherin domains.
CC   -!- SIMILARITY: Contains 4 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF172329; AAF02618.1; -.
DR   EMBL; AB028498; BAA84069.1; -.
DR   EMBL; AE003828; AAF58763.3; -.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0024836; stan.
DR   GO; GO:0016021; C:integral to membrane; NAS.
```

```
DR   GO; GO:0008014; F:calcium-dependent cell adhesion molecule ac. . .; IMP.
DR   GO; GO:0005057; F:receptor signaling protein activity; IMP.
DR   GO; GO:0016358; P:dendrite morphogenesis; IEP.
DR   GO; GO:0007222; P:frizzled signaling pathway; IMP.
DR   GO; GO:0007367; P:segment polarity determination; IMP.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 8.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00053; laminin_EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00112; CA; 8.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00232; CADHERIN_1; 6.
DR   PROSITE; PS50268; CADHERIN_2; 8.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   Cell adhesion; Developmental protein; G-protein coupled receptor;
KW   Calcium-binding; Repeat; Signal; Transmembrane; EGF-like domain;
KW   Laminin EGF-like domain; Glycoprotein.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30   3579       PROTOCADHERIN-LIKE WING POLARITY PROTEIN
FT                                STAN.
FT   DOMAIN       30   2816       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2817   2837       1 (POTENTIAL).
FT   DOMAIN     2838   2845       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2846   2866       2 (POTENTIAL).
FT   DOMAIN     2867   2883       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2884   2904       3 (POTENTIAL).
FT   DOMAIN     2905   2919       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2920   2940       4 (POTENTIAL).
FT   DOMAIN     2941   2959       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2960   2980       5 (POTENTIAL).
FT   DOMAIN     2981   3000       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   3001   3021       6 (POTENTIAL).
FT   DOMAIN     3022   3031       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   3032   3052       7 (POTENTIAL).
FT   DOMAIN     3053   3579       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      360    464       CADHERIN 1.
FT   DOMAIN      465    581       CADHERIN 2.
FT   DOMAIN      582    689       CADHERIN 3.
FT   DOMAIN      690    794       CADHERIN 4.
FT   DOMAIN      795    897       CADHERIN 5.
FT   DOMAIN      898   1007       CADHERIN 6.
FT   DOMAIN     1008   1113       CADHERIN 7.
FT   DOMAIN     1114   1220       CADHERIN 8.
FT   DOMAIN     1482   1518       EGF-LIKE 1 CALCIUM-BINDING.
FT   DOMAIN     1556   1753       LAMININ G-LIKE 1.
FT   DOMAIN     1756   1792       EGF-LIKE 2 CALCIUM-BINDING.
```

```
FT   DOMAIN      1796   1963       LAMININ G-LIKE 2.
FT   DOMAIN      1965   2000       EGF-LIKE 3 CALCIUM-BINDING.
FT   DOMAIN      2091   2126       EGF-LIKE 4 CALCIUM-BINDING.
FT   DOMAIN      2744   2802       GPS.
FT   DOMAIN       140    143       POLY-HIS.
FT   DOMAIN       155    159       POLY-ARG.

  Query Match             14.4%;  Score 225;  DB 1;  Length 3579;
  Best Local Similarity   22.4%;  Pred. No. 4.7e-08;
  Matches   83;  Conservative   52;  Mismatches  120;  Indels  116;  Gaps   17;

Qy          1 CERQPCQHGATCMP------AGEY---------------EFQCLCRDGFKGD----LCEH 35
              | |:|| :   |:       | |:                | | | :|| |     ||:
Db       1422 CVREPCLNFEECLTVLKFGNASEFIHSDTVLFRPIYPVNTFACSCPEGFTGSKEHYLCDT 1481

Qy         36 EENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGH------GIAESDWHLEG 85
              | : |   :|| :||||     |  |:|    :|  |: | ||         |
Db       1482 EVDLC-YSDPCQNGGTCVRREGGYTCVCPSTHTGQNCETGVGHLRPCPSETCEGGLSCLS 1540

Qy         86 SGGNDAPGQYGA-------YFHDDGFLAFPG----HVFSRSLPEVPETIELEVRTSTASG 134
              :  :  |  | |        |  : || |      | |:         ::|   |   :|
Db       1541 NYPSSQPPPYTATCELRARAFGRNSFLTFESLKQRHRFN---------LKLRFATVQENG 1591

Qy        135 LLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARL--VSEDPINDGEWHRVTAL 192
              |||:     |   :  |||:| : :||: | : ||    |:  : |  ::||:||:|  :
Db       1592 LLLYN----GRYNELHDFIALEIHEGHVSFSFSLGDHSERISVIQEAKVSDGKWHQVEVV 1647

Qy        193 REGRRGSIQVDGEE---LVSGR----------------------SPGPNVAVNAKGSVYI 227
                 |  :: :|  :    :||:                      :   :  ::  | : :
Db       1648 YLNRSVTLVLDNCDTAIALSGQLGDRWSCANRTTLKLDKRCSLLTETCHRFLDLTGPLQV 1707

Qy        228 GGAPDVAT---LTGGRFSSGITGCVKNLVL-------------HSARPGAPPPQPLDLQH 271
              || | :     :|   |     ||: :| :             :    | |   ||
Db       1708 GGLPRIPAHFPVTNRDF----VGCISDLRIDDRFVDLNSYVADNGTLAGCPQKAPL---- 1759

Qy        272 RAQAGANTRPC 282
                     : ||
Db       1760 -----CQSEPC 1765


RESULT 11
LMA2_HUMAN
ID   LMA2_HUMAN     STANDARD;      PRT;  3110 AA.
AC   P24043; Q14736; Q93022;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy
DE   chain).
GN   LAMA2 OR LAMM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=94124633; PubMed=8294519;
RA   Vuolteenaho R., Nissinen M., Sainio K., Byers M., Eddy R.,
RA   Hirvonen H., Shows T.B., Sariola H., Engvall E., Tryggvason K.;
RT   "Human laminin M chain (merosin): complete primary structure,
RT   chromosomal assignment, and expression of the M and A chain in human
RT   fetal tissues.";
RL   J. Cell Biol. 124:381-394(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97066955; PubMed=8910357;
RA   Zhang X., Vuolteenaho R., Tryggvason K.;
RT   "Structure of the human laminin alpha2-chain gene (LAMA2), which is
RT   affected in congenital muscular dystrophy.";
RL   J. Biol. Chem. 271:27664-27669(1996).
RN   [3]
RP   SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Placenta;
RX   MEDLINE=90238994; PubMed=2185464;
RA   Ehrig K., Leivo I., Argraves W.S., Ruoslahti E., Engvall E.;
RT   "Merosin, a tissue-specific basement membrane protein, is a
RT   laminin-like protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).
RN   [4]
RP   VARIANTS GLN-545; HIS-619; LEU-919; HIS-2586 AND LYS-2614.
RA   Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA   Marzluf G.A., Amato A.A., Mendell J.R.;
RT   "Novel single base polymorphisms and rare sequence variants in
RT   the laminin 2-chain coding region detected by RNA/SSCP analysis.";
RL   Hum. Mutat. 13:174-174(1999).
RN   [5]
RP   ERRATUM.
RA   Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,
RA   Marzluf G.A., Amato A.A., Mendell J.R.;
RL   Hum. Mutat. 13:340-340(1999).
RN   [6]
RP   VARIANT MDC1A PRO-2564.
RX   MEDLINE=21476011; PubMed=11591858;
RA   He Y., Jones K.J., Vignier N., Morgan G., Chevallay M., Barois A.,
RA   Estournet-Mathiaud B., Hori H., Mizuta T., Tome F.M.S., North K.N.,
RA   Guicheney P.;
RT   "Congenital muscular dystrophy with primary partial laminin alpha-2
RT   chain deficiency: molecular study.";
RL   Neurology 57:1319-1322(2001).
RN   [7]
RP   VARIANTS MDC1A TYR-527 AND ARG-862.
RX   MEDLINE=22439669; PubMed=12552556;
RA   Tezak Z., Prandini P., Boscaro M., Marin A., Devaney J., Marino M.,
RA   Fanin M., Trevisan C.P., Park J., Tyson W., Finkel R., Garcia C.,
RA   Angelini C., Hoffman E.P., Pegoraro E.;
RT   "Clinical and molecular study in congenital muscular dystrophy with
RT   partial laminin alpha-2 (LAMA2) deficiency.";
RL   Hum. Mutat. 21:103-111(2003).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-2 chain is a subunit of laminin-2 (merosin) and
CC       laminin-4 (S-merosin).
CC   -!- SUBCELLULAR LOCATION: Extracellular; found in the basement
CC       membranes (major component).
CC   -!- TISSUE SPECIFICITY: Placenta, striated muscle, peripheral nerve,
CC       cardiac muscle, pancreas, lung, spleen, kidney, adrenal gland,
CC       skin, testis, meninges, choroid plexus, and some other regions of
CC       the brain; not in liver, thymus and bone.
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI, IV and G are globular.
CC   -!- DISEASE: Defects in LAMA2 are the cause of merosin-deficient
CC       congenital muscular dystrophy type 1A (MDC1A) [MIM:607855]. MDC1A
CC       is characterized by difficulty walking, hypotonia, proximal
CC       weakness, hyporeflexia, and white matter hypodensity on MRI.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 17 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z26653; CAA81394.1; -.
```

```
DR   EMBL; U66796; AAB18388.1; -.
DR   EMBL; U66733; AAB18388.1; JOINED.
DR   EMBL; U66734; AAB18388.1; JOINED.
DR   EMBL; U66735; AAB18388.1; JOINED.
DR   EMBL; U66736; AAB18388.1; JOINED.
DR   EMBL; U66737; AAB18388.1; JOINED.
DR   EMBL; U66738; AAB18388.1; JOINED.
DR   EMBL; U66739; AAB18388.1; JOINED.
DR   EMBL; U66740; AAB18388.1; JOINED.
DR   EMBL; U66741; AAB18388.1; JOINED.
DR   EMBL; U66742; AAB18388.1; JOINED.
DR   EMBL; U66743; AAB18388.1; JOINED.
DR   EMBL; U66745; AAB18388.1; JOINED.
DR   EMBL; U66746; AAB18388.1; JOINED.
DR   EMBL; U66747; AAB18388.1; JOINED.
DR   EMBL; U66748; AAB18388.1; JOINED.
DR   EMBL; U66749; AAB18388.1; JOINED.
DR   EMBL; U66750; AAB18388.1; JOINED.
DR   EMBL; U66751; AAB18388.1; JOINED.
DR   EMBL; U66752; AAB18388.1; JOINED.
DR   EMBL; U66753; AAB18388.1; JOINED.
DR   EMBL; U66754; AAB18388.1; JOINED.
DR   EMBL; U66755; AAB18388.1; JOINED.
DR   EMBL; U66756; AAB18388.1; JOINED.
DR   EMBL; U66757; AAB18388.1; JOINED.
DR   EMBL; U66758; AAB18388.1; JOINED.
DR   EMBL; U66759; AAB18388.1; JOINED.
DR   EMBL; U66760; AAB18388.1; JOINED.
DR   EMBL; U66761; AAB18388.1; JOINED.
DR   EMBL; U66762; AAB18388.1; JOINED.
DR   EMBL; U66763; AAB18388.1; JOINED.
DR   EMBL; U66764; AAB18388.1; JOINED.
DR   EMBL; U66765; AAB18388.1; JOINED.
DR   EMBL; U66766; AAB18388.1; JOINED.
DR   EMBL; U66768; AAB18388.1; JOINED.
DR   EMBL; U66769; AAB18388.1; JOINED.
DR   EMBL; U66770; AAB18388.1; JOINED.
DR   EMBL; U66771; AAB18388.1; JOINED.
DR   EMBL; U66772; AAB18388.1; JOINED.
DR   EMBL; U66773; AAB18388.1; JOINED.
DR   EMBL; U66774; AAB18388.1; JOINED.
DR   EMBL; U66775; AAB18388.1; JOINED.
DR   EMBL; U66776; AAB18388.1; JOINED.
DR   EMBL; U66777; AAB18388.1; JOINED.
DR   EMBL; U66778; AAB18388.1; JOINED.
DR   EMBL; U66779; AAB18388.1; JOINED.
DR   EMBL; U66780; AAB18388.1; JOINED.
DR   EMBL; U66781; AAB18388.1; JOINED.
DR   EMBL; U66782; AAB18388.1; JOINED.
DR   EMBL; U66783; AAB18388.1; JOINED.
DR   EMBL; U66784; AAB18388.1; JOINED.
DR   EMBL; U66785; AAB18388.1; JOINED.
DR   EMBL; U66786; AAB18388.1; JOINED.
DR   EMBL; U66787; AAB18388.1; JOINED.
DR   EMBL; U66788; AAB18388.1; JOINED.
DR   EMBL; U66789; AAB18388.1; JOINED.
DR   EMBL; U66790; AAB18388.1; JOINED.
DR   EMBL; U66791; AAB18388.1; JOINED.
DR   EMBL; U66792; AAB18388.1; JOINED.
DR   EMBL; U66793; AAB18388.1; JOINED.
DR   EMBL; U66794; AAB18388.1; JOINED.
DR   EMBL; U66795; AAB18388.1; JOINED.
DR   EMBL; M59832; AAA63215.1; -.
DR   PIR; PX0082; MMHUMH.
DR   HSSP; Q60675; 1QU0.
DR   Genew; HGNC:6482; LAMA2.
DR   MIM; 156225; -.
DR   MIM; 607855; -.
DR   GO; GO:0005604; C:basement membrane; TAS.
DR   GO; GO:0005198; F:structural molecule activity; TAS.
DR   GO; GO:0007517; P:muscle development; TAS.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 5.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD003031; Laminin_B; 1.
DR   SMART; SM00180; EGF_Lam; 15.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 11.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism;
KW   Disease mutation.
FT   SIGNAL        1     22       POTENTIAL.

  Query Match             14.3%;  Score 223.5;  DB 1;  Length 3110;
  Best Local Similarity   36.4%;  Pred. No. 5.1e-08;
  Matches   59;  Conservative   22;  Mismatches   68;  Indels   13;  Gaps    5;

Qy        122 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 181
              ||||||||   |||| :       |    || :: |::|   | | ||||:   :    |
Db       2788 TIELEVRTEAESGLLFYMA-----AINHADFATVQLRNGLPYFSYDLGSGDTHTMIPTKI 2842

Qy        182 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 238
              |||:||::  :|  : | : |||    | |:  |  |  ::  | :|:|| | :  |
Db       2843 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGLPINYTTRRI 2899

Qy        239 GRFSSGITGCVKNLVLHSARPGAPPPQPLDLQHRAQAGANTR 280
              |  :  | |||:|  || |   |   ||    |     || :
Db       2900 GPVTYSIDGCVRN--LHMAEAPADLEQPTSSFHVGTCFANAQ 2939


RESULT 12
NX3A_RAT
ID   NX3A_RAT       STANDARD;      PRT;  1578 AA.
AC   Q07310; Q07280; Q07311; Q07312; Q07313; Q07314;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurexin 3-alpha precursor (Neurexin III-alpha).
GN   NRXN3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=93342001; PubMed=8341647;
RA   Ushkaryov Y.A., Suedhof T.C.;
RT   "Neurexin III alpha: extensive alternative splicing generates
RT   membrane-bound and soluble forms.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:6410-6414(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=95209856; PubMed=7695896;
RA   Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT   "Cartography of neurexins: more than 1000 isoforms generated by
RT   alternative splicing and expressed in distinct subsets of neurons.";
RL   Neuron 14:497-507(1995).
RN   [3]
RP   SEQUENCE OF N-TERMINUS, AND INTERACTION WITH NEUREXOPHILIN 1.
RX   MEDLINE=99074239; PubMed=9856994;
```

```
RA   Missler M., Hammer R.E., Suedhof T.C.;
RT   "Neurexophilin binding to alpha-neurexins. A single LNS domain
RT   functions as an independently folding ligand-binding unit.";
RL   J. Biol. Chem. 273:34716-34723(1998).
RN   [4]
RP   INTERACTION WITH CASK.
RX   MEDLINE=96256685; PubMed=8786425;
RA   Hata Y., Butz S., Suedhof T.C.;
RT   "CASK: a novel dlg/PSD95 homolog with an N-terminal
RT   calmodulin-dependent protein kinase domain identified by interaction
RT   with neurexins.";
RL   J. Neurosci. 16:2488-2494(1996).
RN   [5]
RP   INTERACTION WITH ALPHA-DYSTROGLYCAN.
RX   MEDLINE=21363578; PubMed=11470830;
RA   Sugita S., Saito F., Tang J., Satz J., Campbell K., Suedhof T.C.;
RT   "A stoichiometric complex of neurexins and dystroglycan in brain.";
RL   J. Cell Biol. 154:435-445(2001).
CC   -!- FUNCTION: Neuronal cell surface protein that may be involved in
CC       cell recognition and cell adhesion. May mediate intracellular
CC       signaling.
CC   -!- SUBUNIT: The laminin G-like domain 2 binds to neurexophilin 1.
CC       Isoforms alpha 4B bind to alpha-dystroglycan. The cytoplasmic C-
CC       terminal region binds to CASK.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential). Some
CC       isoforms seems to be secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=1;
CC         Comment=At least 288 isoforms may be produced by alternative
CC         splicing. There is a combination of five alternative spliced
CC         domains at sites 1, 3, 4 and 5, each consisting of modular
CC         sequences (A-L) that seem to be used independently. Isoforms
CC         Alpha 5A-type to Alpha 5H-type lack the transmembrane domain;
CC       Name=Alpha 1A3A4A5I;
CC         IsoId=Q07310-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- SIMILARITY: Contains 6 laminin G-like domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Belongs to the neurexin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L14851; AAA02857.1; -.
DR   EMBL; L14851; AAA02856.1; -.
DR   EMBL; L14851; AAA02858.1; -.
DR   EMBL; L14851; AAA02853.1; -.
DR   EMBL; L14851; AAA02855.1; -.
DR   EMBL; L14851; AAA02854.1; -.
DR   HSSP; Q63373; 1C4R.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003585; Neurexin-like.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 6.
DR   SMART; SM00294; 4.1m; 1.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00282; LamG; 6.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 6.
KW   Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW   Glycoprotein; Alternative splicing.
```

```
FT   SIGNAL        1     27
FT   CHAIN        28   1578       NEUREXIN 3-ALPHA.
FT   DOMAIN       28   1503       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1504   1524       POTENTIAL.
FT   DOMAIN     1525   1578       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       28    202       LAMININ G-LIKE 1.
FT   DOMAIN      198    235       EGF-LIKE 1.
FT   DOMAIN      260    444       LAMININ G-LIKE 2.
FT   DOMAIN      451    643       LAMININ G-LIKE 3.
FT   DOMAIN      647    684       EGF-LIKE 2.
FT   DOMAIN      689    861       LAMININ G-LIKE 4.
FT   DOMAIN      875   1050       LAMININ G-LIKE 5.
FT   DOMAIN     1053   1090       EGF-LIKE 3.
FT   DOMAIN     1094   1294       LAMININ G-LIKE 6.
FT   CARBOHYD     58     58       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    105    105       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    761    761       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1193   1193       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1291   1291       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1335   1335       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1500   1500       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DOMAIN      282    285       POLY-SER.
FT   DOMAIN     1323   1326       POLY-THR.
FT   DOMAIN     1329   1332       POLY-THR.
FT   DOMAIN     1407   1410       POLY-THR.
FT   DOMAIN     1422   1425       POLY-SER.
FT   DOMAIN     1511   1514       POLY-ALA.
FT   VARSPLIC    237    242       Missing (in some isoforms Alpha 1C and
FT                                Alpha 1D).
FT                                /FTId=VSP_003521.
FT   VARSPLIC    237    256       Missing (in some isoforms Alpha 1F).
FT                                /FTId=VSP_003522.
FT   VARSPLIC    243    256       Missing (in some isoforms Alpha 1E).
FT                                /FTId=VSP_003523.
FT   VARSPLIC    253    256       Missing (in some isoforms Alpha 1B and
FT                                Alpha 1D).
FT                                /FTId=VSP_003524.
FT   VARSPLIC    754    763       DCIRINCNSS -> G (in some isoforms Alpha
FT                                3B).
FT                                /FTId=VSP_003525.
FT   VARSPLIC   1210   1239       Missing (in some isoforms Alpha 4B).
FT                                /FTId=VSP_003526.
FT   VARSPLIC   1369   1371       Missing (in some isoforms Alpha 5B, Alpha
FT                                5D, Alpha 5F, Alpha 5H, Alpha 5J and
FT                                Alpha 5L).
FT                                /FTId=VSP_003527.
FT   VARSPLIC   1372   1478       Missing (in some isoforms Alpha 5A, Alpha
FT                                5B, Alpha 5C, Alpha 5D, Alpha 5E, Alpha
FT                                5F, Alpha 5G, Alpha 5H, Alpha 5K and
FT                                Alpha 5L).
FT                                /FTId=VSP_003528.
FT   VARSPLIC   1479   1578       ANPTEPGIRRVPGASEVIRESNSTTGMVVGIVAAAALCILI
FT                                LLYAMYKYRNRDEGSYQVDETRNYISNSAQSNGTLMKEKQA
FT                                SSKSGHKKQKNKDKEYYV -> ARSSNAARITPCRPYMDMA
FT                                THLHIYPSHLHLLCSTLIDTPLPFPHPFFPMLPPSLALLKF
FT                                MCCHPPP (in some isoforms Alpha 5A and
FT                                Alpha 5B).
FT                                /FTId=VSP_003529.
FT   VARSPLIC   1479   1578       ANPTEPGIRRVPGASEVIRESNSTTGMVVGIVAAAALCILI
FT                                LLYAMYKYRNRDEGSYQVDETRNYISNSAQSNGTLMKEKQA
FT                                SSKSGHKKQKNKDKEYYV -> VLERRIILNLKTNAHPKSL
FT                                QSKTC (in some isoforms Alpha 5C and
FT                                Alpha 5D).
FT                                /FTId=VSP_003530.
FT   VARSPLIC   1479   1578       ANPTEPGIRRVPGASEVIRESNSTTGMVVGIVAAAALCILI
FT                                LLYAMYKYRNRDEGSYQVDETRNYISNSAQSNGTLMKEKQA
FT                                SSKSGHKKQKNKDKEYYV -> DILLKSF (in some
FT                                isoforms Alpha 5E and Alpha 5F).
FT                                /FTId=VSP_003531.
FT   VARSPLIC   1479   1578       ANPTEPGIRRVPGASEVIRESNSTTGMVVGIVAAAALCILI
FT                                LLYAMYKYRNRDEGSYQVDETRNYISNSAQSNGTLMKEKQA
FT                                SSKSGHKKQKNKDKEYYV -> ATTTTKKSNFQECGNSICP
FT                                RAFLHNFLL (in some isoforms Alpha 5G and
```

```
FT                                   Alpha 5H).
FT                                   /FTId=VSP_003532.
SQ   SEQUENCE   1578 AA;  173993 MW;  936CF8529143D0C7 CRC64;

  Query Match            13.7%;  Score 214;  DB 1;  Length 1578;
  Best Local Similarity  26.1%;  Pred. No. 1.2e-07;
  Matches   73;  Conservative   38;  Mismatches  121;  Indels   48;  Gaps   10;

Qy         27 GFKGDLCE------------------HEENPCQLREPCLHGGTC----QGTRCLC-LPGF 63
              |||| : :                    | ||  | || :|| |        | |   |:
Db        171 GFKGLMLDLKYGNSEPRLLGSQSVQLEAEGPCGER-PCENGGICFLLDGHPTCDCSTTGY 229

Qy         64 SGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETI 123
               |  | :    |   |   :   | : |  :  | |    :|::   :    :      |
Db        230 GGTLCSEDVSQGPGLSHLMMSEQGRSKAREENVATFRGSEYLSY--DLSQNPIQSSSSEI 287

Qy        124 ELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI-- 181
               |  :|   :||:|         |:  |:::| |:|| :     ||||    : | |:
Db        288 TLSFKTWQRNGLIL-------HTGKSADYVNLALKDGAVSLVINLGSGAFEAIVE-PVNG 339

Qy        182 --NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGG 239
                ||  || |   |  |: :| |||    :| :      : :  | |:| :|  | | |
Db        340 KFNDNAWHDVKVTRNLRQVTISVDGILTTTGYTQEDYTMLGSDDSSYVGPSPSTADLPGS 399

Qy        240 RFSSGITGCVKNLVL--HSARPGAPPPQPLDLQHRAQAGA 277
                |:   ||:| :|   :  |        |:|   |: ||
Db        400 PVSNNFMGCLKEVVYKNNDIR--------LELSRLARIGA 431


RESULT 13
CLR2_HUMAN
ID   CLR2_HUMAN     STANDARD;      PRT;  2923 AA.
AC   Q9HCU4; Q92566;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Epidermal
DE   growth factor-like 2) (Multiple epidermal growth factor-like domains
DE   3) (Flamingo 1).
GN   CELSR2 OR CDHF10 OR EGFL2 OR MEGF3 OR KIAA0279.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20363102; PubMed=10907856;
RA   Vincent J.B., Skaug J., Scherer S.W.;
RT   "The human homologue of flamingo, EGFL2, encodes a brain-expressed
RT   large cadherin-like protein with epidermal growth factor-like domains,
RT   and maps to chromosome 1p13.3-p21.1.";
RL   DNA Res. 7:233-235(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA   Tsutsumi S., Aburatani H., Asai K., Akiyama Y.;
RT   "Genome-wide discovery and analysis of human seven transmembrane helix
RT   receptor genes.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 516-2923 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA   Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
CC   -!- FUNCTION: Receptor that may have an important role in cell/cell
CC       signaling during nervous system formation.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Highest expression in brain and testis.
```

```
CC   -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC   -!- SIMILARITY: Contains 9 cadherin domains.
CC   -!- SIMILARITY: Contains 8 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF234887; AAG00080.1; -.
DR   EMBL; AB065955; BAC06168.1; -.
DR   EMBL; D87469; BAA13407.1; -.
DR   HSSP; P15116; 1NCJ.
DR   Genew; HGNC:3231; CELSR2.
DR   MIM; 604265; -.
DR   GO; GO:0016021; C:integral to membrane; NAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; NAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; NAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 8.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00053; laminin_EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00112; CA; 9.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS00232; CADHERIN_1; 7.
DR   PROSITE; PS50268; CADHERIN_2; 9.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW   Developmental protein; Hydroxylation; Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32   2923       CADHERIN EGF LAG SEVEN-PASS G-TYPE
FT                                RECEPTOR 2.
FT   DOMAIN       32   2380       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2381   2401       1 (POTENTIAL).
FT   DOMAIN     2402   2416       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2417   2437       2 (POTENTIAL).
FT   DOMAIN     2438   2438       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM   2439   2459       3 (POTENTIAL).
FT   DOMAIN     2460   2480       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2481   2501       4 (POTENTIAL).
FT   DOMAIN     2502   2519       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2520   2540       5 (POTENTIAL).
FT   DOMAIN     2541   2560       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   2561   2581       6 (POTENTIAL).
FT   DOMAIN     2582   2591       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2592   2612       7 (POTENTIAL).
FT   DOMAIN     2613   2923       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      182    289       CADHERIN 1.
FT   DOMAIN      290    399       CADHERIN 2.
FT   DOMAIN      400    505       CADHERIN 3.
FT   DOMAIN      506    610       CADHERIN 4.
FT   DOMAIN      611    712       CADHERIN 5.
FT   DOMAIN      713    815       CADHERIN 6.
FT   DOMAIN      816    921       CADHERIN 7.
FT   DOMAIN      922   1023       CADHERIN 8.
FT   DOMAIN     1028   1146       CADHERIN 9.
FT   DOMAIN     1228   1286       EGF-LIKE 1, CALCIUM-BINDING.
FT   DOMAIN     1288   1324       EGF-LIKE 2, CALCIUM-BINDING.
FT   DOMAIN     1328   1366       EGF-LIKE 3, CALCIUM-BINDING.
FT   DOMAIN     1367   1571       LAMININ G-LIKE 1.
FT   DOMAIN     1574   1610       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN     1614   1791       LAMININ G-LIKE 2.
FT   DOMAIN     1793   1828       EGF-LIKE 5, CALCIUM-BINDING.
FT   DOMAIN     1829   1867       EGF-LIKE 6, CALCIUM-BINDING.
FT   DOMAIN     1883   1922       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN     1923   1955       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN     2316   2368       GPS.
FT   DOMAIN     2743   2752       POLY-GLU.
FT   DISULFID   1232   1243       BY SIMILARITY.
FT   DISULFID   1237   1274       BY SIMILARITY.
FT   DISULFID   1276   1285       BY SIMILARITY.
FT   DISULFID   1292   1303       BY SIMILARITY.
FT   DISULFID   1297   1312       BY SIMILARITY.
FT   DISULFID   1314   1323       BY SIMILARITY.
FT   DISULFID   1332   1343       BY SIMILARITY.
FT   DISULFID   1337   1353       BY SIMILARITY.
FT   DISULFID   1355   1365       BY SIMILARITY.
FT   DISULFID   1578   1589       BY SIMILARITY.
FT   DISULFID   1583   1598       BY SIMILARITY.
FT   DISULFID   1600   1609       BY SIMILARITY.
FT   DISULFID   1797   1808       BY SIMILARITY.
FT   DISULFID   1802   1817       BY SIMILARITY.
FT   DISULFID   1819   1828       BY SIMILARITY.
FT   DISULFID   1832   1843       BY SIMILARITY.
FT   DISULFID   1837   1855       BY SIMILARITY.
FT   DISULFID   1857   1866       BY SIMILARITY.
FT   DISULFID   1887   1899       BY SIMILARITY.
FT   DISULFID   1889   1906       BY SIMILARITY.
FT   DISULFID   1908   1921       BY SIMILARITY.
FT   DISULFID   1924   1936       BY SIMILARITY.
FT   DISULFID   1926   1943       BY SIMILARITY.
FT   DISULFID   1945   1954       BY SIMILARITY.
FT   MOD_RES    1591   1591       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1810   1810       HYDROXYLATION (POTENTIAL).
FT   CARBOHYD    486    486       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    557    557       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    701    701       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1036   1036       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1076   1076       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1182   1182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1212   1212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1501   1501       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1565   1565       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1741   1741       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1827   1827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1900   1900       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2024   2024       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2043   2043       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2061   2061       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2323   2323       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   2345   2345       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2923 AA;  317447 MW;  382757D315158ED8 CRC64;

  Query Match             13.5%;  Score 211.5;  DB 1;  Length 2923;
  Best Local Similarity   23.3%;  Pred. No. 3.4e-07;
  Matches   81;  Conservative   43;  Mismatches  132;  Indels   91;  Gaps   15;

Qy          1 CERQPCQHGATCMPAGEYE---------------------FQCLCRDGFKGDLCEHEENP 39
              | |:||::   |:    ::                      :| |  || || || | :
Db       1232 CLREPCENYMRCVSVLRFDSSAPFIASSSVLFRPIHPVGGLRCRCPPGFTGDYCETEVDL 1291

Qy         40 CQLREPCLHGGTCQ----GTRCLCLPGFSGPRCQQG--SGH---GIAESDWH----LEGS 86
              |  | ||   | |:    |  |||  |::|  |:    ||    |: ::       | |
Db       1292 CYSR-PCGPHGRCRSREGGYTCLCRDGYTGEHCEVSARSGRCTPGVCKNGGTCVNLLVGG 1350

Qy         87 GGNDAP-GQYGAYFHDDGFLAFPGHVF--SRSLPE-VPETIELEVRTSTASGLLLWQGVE 142
                 | | | :   :      :|| | |   | | :    |: |   |    ||||:
Db       1351 FKCDCPSGDFEKPYCQVTTRSFPAHSFITFRGLRQRFHFTLALSFATKERDGLLLYN--- 1407

Qy        143 VGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP--INDGEWHRV----------- 189
               |   :  ||::| :    :   :  |     :    |  ::||:|| |
Db       1408 -GRFNEKHDFVALEVIQEQVQLTFSAGESTTTVSPFVPGGVSDGQWHTVQLKYYNKPLLG 1466

Qy        190 -TALREG--------------------RRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIG 228
               | | :|                    | ||:  :      |   |   :::  | : :|
Db       1467 QTGLPQGPSEQKVAVVTVDGCDTGVALRFGSVLGNYSCAAQGTQGGSKKSLDLTGPLLLG 1526

Qy        229 GAPDVATLTGGRFSSGITGCVKNLVLHSAR-------------PGAP 262
              | ||:      |      ||::|| : |               || |
Db       1527 GVPDLPESFPVRMRQ-FVGCMRNLQVDSRHIDMADFIANNGTVPGCP 1572


RESULT 14
CLR2_RAT
ID   CLR2_RAT       STANDARD;      PRT;  2144 AA.
AC   Q9QYP2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cadherin EGF LAG seven-pass G-type receptor 2 (Multiple epidermal
DE   growth factor-like domains 3) (Fragment).
GN   CELSR2 OR MEGF3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98360089; PubMed=9693030;
RA   Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT   "Identification of high-molecular-weight proteins with multiple
RT   EGF-like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
CC   -!- FUNCTION: Receptor that may have an important role in cell/cell
CC       signaling during nervous system formation.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in the brain. High expression in
CC       cerebrellum and olfactory bulb. Weaker expression in cerebral
CC       cortex, hippocampus and brain stem.
CC   -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC   -!- SIMILARITY: Contains 4 cadherin domains.
CC   -!- SIMILARITY: Contains 8 EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin G-like domains.
CC   -!- SIMILARITY: Contains 1 laminin EGF-like domain.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB011529; BAA88687.1; -.
DR   HSSP; P00740; 1EDM.
DR   GO; GO:0016021; C:integral to membrane; ISS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; ISS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; ISS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 3.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00053; laminin_EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00112; CA; 3.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 4.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW   Developmental protein; Hydroxylation.
FT   NON_TER       1      1
FT   DOMAIN        1   1605       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1606   1626       1 (POTENTIAL).
FT   DOMAIN     1627   1641       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1642   1662       2 (POTENTIAL).
FT   DOMAIN     1663   1663       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1664   1684       3 (POTENTIAL).
FT   DOMAIN     1685   1705       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1706   1726       4 (POTENTIAL).
FT   DOMAIN     1727   1744       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1745   1765       5 (POTENTIAL).
FT   DOMAIN     1766   1789       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1790   1810       6 (POTENTIAL).
FT   DOMAIN     1811   1816       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1817   1837       7 (POTENTIAL).
FT   DOMAIN     1838   2144       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       <1     40       CADHERIN 1.
FT   DOMAIN       41    146       CADHERIN 2.
FT   DOMAIN      147    248       CADHERIN 3.
FT   DOMAIN      253    371       CADHERIN 4.
FT   DOMAIN      453    511       EGF-LIKE 1, CALCIUM-BINDING.
FT   DOMAIN      513    549       EGF-LIKE 2, CALCIUM-BINDING.
FT   DOMAIN      553    591       EGF-LIKE 3, CALCIUM-BINDING.
FT   DOMAIN      592    796       LAMININ G-LIKE 1.
```

```
FT   DOMAIN      799    835       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      839   1016       LAMININ G-LIKE 2.
FT   DOMAIN     1018   1053       EGF-LIKE 5, CALCIUM-BINDING.
FT   DOMAIN     1054   1092       EGF-LIKE 6, CALCIUM-BINDING.
FT   DOMAIN     1108   1147       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN     1148   1180       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN     1541   1593       GPS.
FT   DOMAIN     1968   1973       POLY-GLU.
FT   DISULFID    457    468       BY SIMILARITY.
FT   DISULFID    462    499       BY SIMILARITY.
FT   DISULFID    501    510       BY SIMILARITY.
FT   DISULFID    517    528       BY SIMILARITY.
FT   DISULFID    522    537       BY SIMILARITY.
FT   DISULFID    539    548       BY SIMILARITY.
FT   DISULFID    557    568       BY SIMILARITY.
FT   DISULFID    562    578       BY SIMILARITY.
FT   DISULFID    580    590       BY SIMILARITY.
FT   DISULFID    803    814       BY SIMILARITY.
FT   DISULFID    808    823       BY SIMILARITY.
FT   DISULFID    825    834       BY SIMILARITY.
FT   DISULFID   1022   1033       BY SIMILARITY.
FT   DISULFID   1027   1042       BY SIMILARITY.
FT   DISULFID   1044   1053       BY SIMILARITY.
FT   DISULFID   1057   1068       BY SIMILARITY.
FT   DISULFID   1062   1080       BY SIMILARITY.
FT   DISULFID   1082   1091       BY SIMILARITY.
FT   DISULFID   1112   1124       BY SIMILARITY.
FT   DISULFID   1114   1131       BY SIMILARITY.
FT   DISULFID   1133   1146       BY SIMILARITY.
FT   DISULFID   1149   1161       BY SIMILARITY.
FT   DISULFID   1151   1168       BY SIMILARITY.
FT   DISULFID   1170   1179       BY SIMILARITY.
FT   MOD_RES     816    816       HYDROXYLATION (POTENTIAL).
FT   MOD_RES    1035   1035       HYDROXYLATION (POTENTIAL).
FT   CARBOHYD    261    261       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    407    407       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    437    437       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    726    726       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    790    790       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    966    966       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1052   1052       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1125   1125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1249   1249       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1268   1268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1286   1286       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1548   1548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1570   1570       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2144 AA;  233480 MW;  6EA898C1BA655ECA CRC64;

  Query Match             13.3%;  Score 207.5;  DB 1;  Length 2144;
  Best Local Similarity   23.6%;  Pred. No. 4.7e-07;
  Matches   82;  Conservative   42;  Mismatches  132;  Indels   91;  Gaps   16;

Qy          1 CERQPCQHGATCMPAGEYE----------------------FQCLCRDGFKGDLCEHEENP 39
              | |:||::   |:    ::                      :| |  || || || | :
Db        457 CLREPCENYMRCVSVLRFDSSAPFIASSSVLFRPIHPVGGLRCRCPPGFTGDYCETEVDL 516

Qy         40 CQLREPCLHGGTCQ----GTRCLCLPGFSGPRCQQG--SGH---GIAESDWH----LEGS 86
              |  | ||   | |:    |  |||  |::|  |:    ||    |: ::       | |
Db        517 CYSR-PCGPHGHCRSREGGYTCLCRDGYTGEHCEVSARSGRCTPGVCKNGGTCVNLLVGG 575

Qy         87 GGNDAP-GQYGAYFHDDGFLAFPGHVF--SRSLPE-VPETIELEVRTSTASGLLLWQGVE 142
                 | | | :   |      :||   |   | | :    |: |   |    ||||:
Db        576 FKCDCPSGDFEKPFCQVTTRSFPARSFITFRGLRQRFHFTLALSFATKERDGLLLYN--- 632

Qy        143 VGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDP--INDGEWHRV----------- 189
               |   :  ||::| :    :   :  |     :    |  ::||:|| |
Db        633 -GRFNEKHDFVALEVIQEQVQLTFSAGESTTTVSPFVPGGVSDGQWHTVQLKYYNKPLLG 691

Qy        190 -TALREG----RRGSIQVDGEE----------------LVSGRSPGPNVAVNAKGSVYIG 228
               | | :|    :   : ||| :                   |   |   :::  | : :|
Db        692 QTGLPQGPSEQKVAVVSVDGCDTGVALRFGAMLGNYSCAAQGTQGGSKKSLDLTGPLLLG 751
```

```
Qy        229 GAPDVATLTGGRFSSGITGCVKNLVLHSAR-------------PGAP 262
              | ||:      |      ||:||| : |               || |
Db        752 GVPDLPESFPVRMRH-FVGCMKNLQVDSRHVDMADFIANNGTVPGCP 797

RESULT 15
LMA2_MOUSE
ID   LMA2_MOUSE     STANDARD;      PRT;  3106 AA.
AC   Q60675; Q05003; Q64061;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy
DE   chain).
GN   LAMA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Embryo, and Heart;
RX   MEDLINE=95316259; PubMed=7795883;
RA   Bernier S.M., Utani A., Sugiyama S., Doi T., Polistina C.,
RA   Yamada Y.;
RT   "Cloning and expression of laminin alpha 2 chain (M-chain) in the
RT   mouse.";
RL   Matrix Biol. 14:447-455(1995).
RN   [2]
RP   SEQUENCE OF 2162-2279 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Thymus;
RX   MEDLINE=93346725; PubMed=8345183;
RA   Chang A.C., Wadsworth S., Coligan J.E.;
RT   "Expression of merosin in the thymus and its interaction with
RT   thymocytes.";
RL   J. Immunol. 151:1789-1801(1993).
RN   [3]
RP   SEQUENCE OF 64-281 FROM N.A.
RX   MEDLINE=95179178; PubMed=7874173;
RA   Xu H., Wu X.R., Wewer U.M., Engvall E.;
RT   "Murine muscular dystrophy caused by a mutation in the laminin alpha
RT   2 (Lama2) gene.";
RL   Nat. Genet. 8:297-302(1994).
RN   [4]
RP   SEQUENCE OF 20-25.
RX   MEDLINE=21818471; PubMed=11829758;
RA   Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
RT   "Complete sequence, recombinant analysis and binding to laminins and
RT   sulphated ligands of the N-terminal domains of laminin alpha3B and
RT   alpha5 chains.";
RL   Biochem. J. 362:213-221(2002).
RN   [5]
RP   BINDING TO FBLN1, FBLN2, AND NID2.
RX   MEDLINE=99146904; PubMed=10022829;
RA   Talts J.F., Andac Z., Goehring W., Brancaccio A., Timpl R.;
RT   "Binding of the G domains of laminin alpha1 and alpha2 chains and
RT   perlecan to heparin, sulfatides, alpha-dystroglycan and several
RT   extracellular matrix proteins.";
RL   EMBO J. 18:863-870(1999).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
RX   MEDLINE=20085745; PubMed=10619025;
RA   Hohenester E., Tisi D., Talts J.F., Timpl R.;
RT   "The crystal structure of a laminin G-like module reveals the
RT   molecular basis of alpha-dystroglycan binding to laminins, perlecan,
RT   and agrin.";
RL   Mol. Cell 4:783-792(1999).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The alpha-2 chain is a subunit of laminin-2 (Merosin) and
CC       laminin-4 (S-merosin). Interacts with FBLN1, FBLN2 and NID2.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes (major
CC       component).
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI, IV and G are globular.
CC   -!- DISEASE: Defects in LAMA2 are a cause of murine muscular dystrophy
CC       (dy2J).
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 17 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 2 laminin IV domains.
CC   -!- SIMILARITY: Contains 5 laminin G-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U12147; AAC52165.1; -.
DR   EMBL; X69869; CAA49502.1; -.
DR   EMBL; S75315; AAB33573.1; -.
DR   PIR; I49077; S53868.
DR   PDB; 1QU0; 03-DEC-99.
DR   PDB; 1DYK; 04-FEB-01.
DR   MGD; MGI:99912; Lama2.
DR   GO; GO:0005604; C:basement membrane; IDA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 2.
DR   Pfam; PF00053; laminin_EGF; 14.
DR   Pfam; PF00054; laminin_G; 5.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD003031; Laminin_B; 1.
DR   SMART; SM00180; EGF_Lam; 14.
DR   SMART; SM00281; LamB; 2.
DR   SMART; SM00282; LamG; 5.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 11.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
FT   SIGNAL        1     19
FT   CHAIN        20   3106       LAMININ ALPHA-2 CHAIN.
FT   DOMAIN       20    282       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      283    339       LAMININ EGF-LIKE 1.
FT   DOMAIN      340    409       LAMININ EGF-LIKE 2.
FT   DOMAIN      410    464       LAMININ EGF-LIKE 3.
FT   DOMAIN      465    513       LAMININ EGF-LIKE 4.
FT   DOMAIN      514    523       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      524    719       LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT   DOMAIN      720    752       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN      753    802       LAMININ EGF-LIKE 6.
FT   DOMAIN      803    860       LAMININ EGF-LIKE 7.
FT   DOMAIN      861    913       LAMININ EGF-LIKE 8.
FT   DOMAIN      914    962       LAMININ EGF-LIKE 9.
FT   DOMAIN      963   1009       LAMININ EGF-LIKE 10.
FT   DOMAIN     1010   1055       LAMININ EGF-LIKE 11.
```

```
FT   DOMAIN     1056   1101       LAMININ EGF-LIKE 12.
FT   DOMAIN     1102   1161       LAMININ EGF-LIKE 13.
FT   DOMAIN     1162   1171       LAMININ EGF-LIKE 14 (N-TERMINAL).
FT   DOMAIN     1172   1375       LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT   DOMAIN     1376   1415       LAMININ EGF-LIKE 14 (C-TERMINAL).
FT   DOMAIN     1416   1464       LAMININ EGF-LIKE 15.
FT   DOMAIN     1465   1522       LAMININ EGF-LIKE 16.
FT   DOMAIN     1523   1569       LAMININ EGF-LIKE 17.
FT   DOMAIN     1570   2140       DOMAIN II AND I.
FT   DOMAIN     2141   2324       LAMININ G-LIKE 1.
FT   DOMAIN     2336   2517       LAMININ G-LIKE 2.
FT   DOMAIN     2522   2706       LAMININ G-LIKE 3.
FT   DOMAIN     2759   2930       LAMININ G-LIKE 4.
FT   DOMAIN     2929   3106       LAMININ G-LIKE 5.
FT   DOMAIN     1662   1863       COILED COIL (POTENTIAL).
FT   DOMAIN     1923   2146       COILED COIL (POTENTIAL).
FT   DISULFID    283    292       BY SIMILARITY.
FT   DISULFID    285    303       BY SIMILARITY.
FT   DISULFID    305    314       BY SIMILARITY.
FT   DISULFID    317    337       BY SIMILARITY.
FT   DISULFID    340    349       BY SIMILARITY.
FT   DISULFID    342    374       BY SIMILARITY.
FT   DISULFID    377    386       BY SIMILARITY.
FT   DISULFID    389    407       BY SIMILARITY.
FT   DISULFID    410    422       BY SIMILARITY.
FT   DISULFID    412    438       BY SIMILARITY.
FT   DISULFID    440    449       BY SIMILARITY.
FT   DISULFID    452    462       BY SIMILARITY.
FT   DISULFID    465    478       BY SIMILARITY.
FT   DISULFID    467    482       BY SIMILARITY.
FT   DISULFID    484    493       BY SIMILARITY.
FT   DISULFID    496    511       BY SIMILARITY.
FT   DISULFID    753    762       BY SIMILARITY.
FT   DISULFID    755    769       BY SIMILARITY.
FT   DISULFID    772    781       BY SIMILARITY.
FT   DISULFID    784    800       BY SIMILARITY.
FT   DISULFID    803    818       BY SIMILARITY.
FT   DISULFID    805    828       BY SIMILARITY.
FT   DISULFID    831    840       BY SIMILARITY.
FT   DISULFID    843    858       BY SIMILARITY.
FT   DISULFID    861    875       BY SIMILARITY.
FT   DISULFID    863    882       BY SIMILARITY.
FT   DISULFID    885    894       BY SIMILARITY.
FT   DISULFID    897    911       BY SIMILARITY.
FT   DISULFID    914    926       BY SIMILARITY.
FT   DISULFID    916    933       BY SIMILARITY.
FT   DISULFID    935    944       BY SIMILARITY.
FT   DISULFID    947    960       BY SIMILARITY.
FT   DISULFID    963    975       BY SIMILARITY.
FT   DISULFID    965    981       BY SIMILARITY.
FT   DISULFID    983    992       BY SIMILARITY.
FT   DISULFID    995   1007       BY SIMILARITY.
FT   DISULFID   1010   1019       BY SIMILARITY.
FT   DISULFID   1012   1026       BY SIMILARITY.
FT   DISULFID   1028   1037       BY SIMILARITY.
FT   DISULFID   1040   1053       BY SIMILARITY.
FT   DISULFID   1056   1068       BY SIMILARITY.
FT   DISULFID   1058   1075       BY SIMILARITY.
FT   DISULFID   1077   1086       BY SIMILARITY.
FT   DISULFID   1089   1099       BY SIMILARITY.
FT   DISULFID   1416   1425       BY SIMILARITY.
FT   DISULFID   1418   1432       BY SIMILARITY.
  Query Match             13.3%;  Score 207.5;  DB 1;  Length 3106;
  Best Local Similarity   35.8%;  Pred. No. 7e-07;
  Matches   54;  Conservative   25;  Mismatches   53;  Indels   19;  Gaps    6;

Qy        122 TIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPI 181
              ||||||||   |||| : |  :  |    || :: |::|   | | ||||  | :    |
Db       2784 TIELEVRTEAESGLLFYMG-RINHA----DFGTVQLRNGFPFFSYDLGSGSTRTMIPTKI 2838

Qy        182 NDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVA--VNAKGSVYIGGAP-DVATLTG 238
              |||:||::  :|  : | : ||     | ::  |  |  ::  | :|:|| | :  |
```

```
Db       2839 NDGQWHKIKIVRVKQEGILYVDD---ASSQTISPKKADILDVGGILYVGGLPINYTTRRI 2895

Qy        239 GRFSSGITGCVKNLVLHSARPGAPPPQPLDL 269
              |  :  : |||:|| :  |        |:||
Db       2896 GPVTYSLDGCVRNLHMEQA--------PVDL 2918

Search completed: March  9, 2004, 17:19:44
Job time : 7.86906 secs
```

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         March  9, 2004, 17:15:08 ; Search time 24.8707 Seconds
                                           (without alignments)
                                           3602.917 Million cell updates/sec

Title:          US-10-006-011A-9
Perfect score:  1566
Sequence:       1 CERQPCQHGATCMPAGEYEF..........QPLDLQHRAQAGANTRPCPS 284

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_25:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                            Description
------------------------------------------------------------------------------
    1   357.5   22.8   2026  4  O00468                         O00468 homo sapien
    2     349   22.3    925  5  Q9U7E8                         Q9u7e8 caenorhabdi
    3   346.5   22.1    488  4  Q96IC1                         Q96ic1 homo sapien
    4     343   21.9    463  4  Q8NAL2                         Q8nal2 homo sapien
    5     343   21.9    775  4  Q8N7Y0                         Q8n7y0 homo sapien
    6     325   20.8    295  11  Q80WX4                         Q80wx4 mus musculu
    7     325   20.8   1009  11  Q8BGP3                         Q8bgp3 mus musculu
    8     304   19.4    263  11  Q80V56                         Q80v56 mus musculu
    9     304   19.4    263  11  Q99KT4                         Q99kt4 mus musculu
   10     304   19.4    350  11  Q8K326                         Q8k326 mus musculu
   11     295   18.8    478  5  Q86SD6                         Q86sd6 ciona intes
   12     292   18.6   1361  5  Q9V7I4                         Q9v7i4 drosophila
   13     292   18.6   1361  5  Q9NGV2                         Q9ngv2 drosophila
   14     292   18.6   1534  4  O75093                         O75093 homo sapien
   15     292   18.6   1618  4  Q9UIL7                         Q9uil7 homo sapien
   16   283.5   18.1   1531  11  O88279                         O88279 rattus norv
   17     281   17.9    708  5  Q9NFS9                         Q9nfs9 drosophila
   18     281   17.9   1035  5  Q9NEG1                         Q9negl drosophila
   19     281   17.9   4117  5  Q8IRV9                         Q8irv9 drosophila
   20     281   17.9   4179  5  Q9W4Y4                         Q9w4y4 drosophila
   21     281   17.9   4223  5  Q8MPN3                         Q8mpn3 drosophila
   22     281   17.9   4228  5  Q8IRV8                         Q8irv8 drosophila
   23   280.5   17.9    738  13  Q90Z45                         Q90z45 gallus gall
   24   280.5   17.9   1458  11  Q8CJG8                         Q8cjg8 rattus norv
   25   280.5   17.9   1531  11  Q9WVB5                         Q9wvb5 mus musculu
   26   280.5   17.9   1557  11  Q80TR4                         Q80tr4 mus musculu
   27   279.5   17.8     68  11  O08591                         O08591 rattus norv
   28   278.5   17.8    152  4  Q8N124                         Q8n124 homo sapien
   29     276   17.6    152  4  Q8N197                         Q8n197 homo sapien
   30   274.5   17.5   1530  11  Q9WUG5                         Q9wug5 rattus norv
   31   270.5   17.3   1474  11  Q8CJG9                         Q8cjg9 rattus norv
   32     265   16.9    181  4  Q8N4J5                         Q8n4j5 homo sapien
   33     264   16.9   4307  5  Q19319                         Q19319 caenorhabdi
   34   258.5   16.5   1529  13  Q7ZXI2                         Q7zxi2 xenopus lae
   35   250.5   16.0   1039  5  Q9VQB1                         Q9vqb1 drosophila
   36     249   15.9   1410  5  Q20204                         Q20204 caenorhabdi
   37   246.5   15.7   1515  13  Q9DE37                         Q9de37 brachydanio
   38     246   15.7   1523  11  O88280                         O88280 rattus norv
   39   244.5   15.6   1512  13  Q9DE36                         Q9de36 brachydanio
   40     244   15.6   1530  13  Q90WZ3                         Q90wz3 xenopus lae
   41   243.5   15.5   1025  11  Q9Z166                         Q9z166 mus musculu
   42   243.5   15.5   1521  11  Q9R1B9                         Q9r1b9 mus musculu
   43   241.5   15.4    783  13  Q90XG2                         Q90xg2 gallus gall
   44   241.5   15.4    950  13  Q90Z44                         Q90z44 gallus gall
   45     241   15.4   1523  4  O75094                         O75094 homo sapien
```

```
                         ALIGNMENTS

RESULT 1
O00468
ID   O00468      PRELIMINARY;      PRT;  2026 AA.
AC   O00468;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   AGRIN precursor (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,
RA   Veerkamp J.H., Monnens L.A.H., van den Heuvel L.P.W.J.;
RL   Eur. J. Biochem. 0:0-0(1998).
RN   [2]
RP   SEQUENCE OF 1-153 FROM N.A.
RX   MEDLINE=96224170; PubMed=8617505;
RA   Lennon G., Auffray C., Polymeropoulos M., Soares M.B.;
RT   "The I.M.A.G.E. Consortium: an integrated molecular analysis of
RT   genomes and their expression.";
RL   Genomics 33:151-152(1996).
RN   [3]
RP   SEQUENCE OF 1-153 FROM N.A.
RA   Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;
RL   J. Cell Biol. 0:0-0(0).
DR   EMBL; AF016903; AAC39776.1; -.
DR   EMBL; U84406; AAB52917.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:329; AGRN.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR004850; Agrin_NtA.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003645; FolN.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
```

```
DR   InterPro; IPR000082; SEA_domain.
DR   InterPro; IPR008993; TIMP_like.
DR   InterPro; IPR001455; UPF0033.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00050; kazal; 9.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF03146; NtA; 1.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 2.
DR   SMART; SM00274; FOLN; 5.
DR   SMART; SM00280; KAZAL; 9.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS50024; SEA; 1.
DR   PROSITE; PS01148; UPF0033; 1.
KW   EGF-like domain; Laminin EGF-like domain; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     10       POTENTIAL.
FT   CHAIN        11   2026       AGRIN.
SQ   SEQUENCE   2026 AA;  212881 MW;  4AB0EE710CD4B8EF CRC64;

  Query Match             22.8%;  Score 357.5;  DB 4;  Length 2026;
  Best Local Similarity   35.4%;  Pred. No. 4.7e-19;
  Matches   91;  Conservative   29;  Mismatches  110;  Indels   27;  Gaps    8;

Qy          1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GT 55
              |   ||  || |       | | |  |  |  |  |::|||   ||     |:     |
Db       1534 CLPNPCHGGAPCQNLEAGRFHCQCPPGRVGPTCADEKSPCQ-PNPCHGAAPCRVLPEGGA 1592

Qy         56 RCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSR 114
              :| |  |  |  ||  |               | |  | : | |:    |   | | |:|
Db       1593 QCECPLGREGTFCQTAS---------------GQDGSGPFLADFNGFSHLELRGLHTFAR 1637

Qy        115 SLPEVPETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEAR 174
               | |    :|:       |||||: | :    |:| ||:|| |:|  | ||| || | |
Db       1638 DLGE-KMALEVVFLARGPSGLLLYNGQKTD--GKG-DFVSLALRDRRLEFRYDLGKGAAV 1693

Qy        175 LVSEDPINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVA 234
              : | :|:  | | ||:  | ||:|:::|     | | || |:  :| |  :|:||||| :
Db       1694 IRSREPVTLGAWTRVSLERNGRKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFS 1753

Qy        235 TLT-GGRFSSGITGCVK 250
               |      |||  | ::
Db       1754 KLARAAAVSSGFDGAIQ 1770

RESULT 2
Q9U7E8
ID   Q9U7E8       PRELIMINARY;      PRT;   925 AA.
AC   Q9U7E8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   UNC-52/perlecan (Fragment).
GN   UNC-52.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE=99443780; PubMed=10512861;
RA   Mullen G.P., Rogalski T.M., Bush J.A., Gorji P.R., Moerman D.G.;
RT   "Complex patterns of alternative splicing mediate the spatial and
RT   temporal distribution of Perlecan/UNC-52 in caenorhabditis elegans.";
RL   Mol. Biol. Cell 10:3205-3221(1999).
```

```
DR   EMBL; AF132883; AAD25092.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS50835; IG_LIKE; 1.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain; Immunoglobulin domain.
FT   NON_TER       1      1
SQ   SEQUENCE   925 AA;  102174 MW;  474BB5F045D67E0B CRC64;

  Query Match             22.3%;  Score 349;  DB 5;  Length 925;
  Best Local Similarity   31.2%;  Pred. No. 8.1e-19;
  Matches   81;  Conservative   41;  Mismatches  102;  Indels   36;  Gaps    6;

Qy         30 GDLCEHEE--NPCQLREPCLHGGTCQGTR-----CLCLPGFSGPRCQQGSGHGIAESDWH 82
              ||:   :|  | |     |   | |         | |   : || |
Db        679 GDVYSTQEPNNIC-ANSTCGMNGQCVPRNMTHYTCECKLYYDGPTCSLFK---------- 727

Qy         83 LEGSGGNDAPGQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVE 142
                       | :: | |  | |:      |     |  | :  : :|   :|:||||| :
Db        728 ---------PIEHAARFDGDAFIELSSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-Q 777

Qy        143 VGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQV 202
                   | :|:||:|: :||| | |:|| | | |:||: ::||: | |   |:|| | :::
Db        778 RPTVQQMEDYISVGIVNGHLHFSYELGGGAAHLISEERVDDGKEHSVRFERKGREGQMRI 837

Qy        203 DGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAP 262
              |    | ||| |    :|  |::::|| ||::  ||| ||:   ||: :: |:  :
Db        838 DNYREVDGRSTGILAMLNVDGNIFVGGVPDISKATGGLFSNNFVGCIADVELNGVK---- 893

Qy        263 PPQPLDLQHRAQAGANTRPC 282
                  |||   |  | | :||
Db        894 ----LDLMATAIDGKNVKPC 909

RESULT 3
Q96IC1
ID   Q96IC1       PRELIMINARY;      PRT;   488 AA.
AC   Q96IC1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007649; AAH07649.1; -.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
```

```
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein; EGF-like domain.
FT   NON_TER       1      1
SQ   SEQUENCE   488 AA;  51764 MW;  A732B99CC680E772 CRC64;

  Query Match             22.1%;  Score 346.5;  DB 4;  Length 488;
  Best Local Similarity   35.3%;  Pred. No. 5.6e-19;
  Matches   89;  Conservative   29;  Mismatches  107;  Indels   27;  Gaps    8;

Qy          6 CQHGATCMPAGEYEFQCLCRDGFKGDLCEHEENPCQLREPCLHGGTCQ-----GTRCLCL 60
              |  || |       | | |  |  |  |  |::|||   ||     |:     | :| |
Db          1 CHGGAPCQNLEAGRFHCQCPPGRVGPTCADEKSPCQ-PNPCHGAAPCRVLPEGGAQCECP 59

Qy         61 PGFSGPRCQQGSGHGIAESDWHLEGSGGNDAPGQYGAYFHDDGFLAFPG-HVFSRSLPEV 119
               |  |  ||  |               | |  | : | |:    |   | | |:| | |
Db         60 LGREGTFCQTAS---------------GQDGSGPFLADFNGFSHLELRGLHTFARDLGE- 103

Qy        120 PETIELEVRTSTASGLLLWQGVEVGEAGQGKDFISLGLQDGHLVFRYQLGSGEARLVSED 179
                 :|:       |||||: | :    |:| ||:|| |:|  | ||| || | | : | :
Db        104 KMALEVVFLARGPSGLLLYNGQKTD--GKG-DFVSLALRDRRLEFRYDLGKGAAVIRSRE 160

Qy        180 PINDGEWHRVTALREGRRGSIQVDGEELVSGRSPGPNVAVNAKGSVYIGGAPDVATLT-G 238
              |:  | | ||:  | ||:|:::|     | | || |:  :| |  :|:||||| : |
Db        161 PVTLGAWTRVSLERNGRKGALRVGDGPRVLGESPVPHTVLNLKEPLYVGGAPDFSKLARA 220

Qy        239 GRFSSGITGCVK 250
                 |||  | ::
Db        221 AAVSSGFDGAIQ 232

RESULT 4
Q8NAL2
ID   Q8NAL2      PRELIMINARY;      PRT;   463 AA.
AC   Q8NAL2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ35160.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK092479; BAC03900.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 2.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   463 AA;  50635 MW;  F688BF2714D5D0EC CRC64;

  Query Match             21.9%;  Score 343;  DB 4;  Length 463;
  Best Local Similarity   33.5%;  Pred. No. 9.9e-19;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db        236 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 281

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
Db        282 EIPQFIGRSYLTYDNPDILKRVSG---SRS------NVFMRFKTTAKDGLLLWRG--DSP 330

Qy        146 AGQGKDFISLGLQDGHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
                   |||||||:|| ||| | |||| | ::     ||| |||| |:|:|: | | ||
Db        331 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 390

Qy        206 ELVSGRSPGPNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKNLVLHSARPGAPPPQ 265
                 :|:|||    :|  |::|:||  ::|  |  ::  |: ||: :  | :
Db        391 GARTGKSPGMMRQLNINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-------DY 443

Qy        266 PLDLQHRAQAGANTRPC 282
               : |   |  | |   |
Db        444 HISLVEDAVDGKNINTC 460

RESULT 5
Q8N7Y0
ID   Q8N7Y0      PRELIMINARY;      PRT;   775 AA.
AC   Q8N7Y0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ40230.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK097549; BAC05096.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   775 AA;  84800 MW;  61A81294F0204ED2 CRC64;

  Query Match             21.9%;  Score 343;  DB 4;  Length 775;
  Best Local Similarity   33.5%;  Pred. No. 1.9e-18;
  Matches   86;  Conservative   34;  Mismatches   97;  Indels   40;  Gaps    8;

Qy         34 EHEENPCQLREPCLHGGTC----QGTRCLCLPGFSGPRCQQGSGHGIAESDWHLEGSGGN 89
              |:  :|| :| || |||:|    :|  | |  || |  ||:     |
Db        548 ENAAHPC-VRAPCAHGGSCRPRKEGYDCDCPLGFEGLHCQKAIIEAI------------- 593

Qy         90 DAP----GQYGAYFHDDGFLAFPGHVFSRSLPEVPETIELEVRTSTASGLLLWQGVEVGE 145
              : |      |  | : |      |   |||       : :  :|:   |||||:|
```